View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by DSpace@MIT

CEEPR Center for Energy and Environmental Policy Research

Transmission Rights and Market Power on Electric Power Networks

by
Paul L. Joskow and Jean Tirole

00-002 WP January 2000

A Joint Center of the Department of Economics, Laboratory for Energy and the Environment, and Sloan School of Management

Transmission Rights and Market Power on Electric Power Networks*

Paul L. Joskow †and Jean Tirole‡

January 27, 2000

Abstract

We analyze whether and how the allocation of transmission rights associated with the use of electric power networks affects the behavior of electricity generators and electricity consumers with market power. We consider two alternative types of transmission rights: (1) financial rights which give the owner a share of the congestion charges which accrue to the network operator when the network is congested and (2) physical rights which give the owner the right to utilize scarce transmission capacity without paying additional congestion charges. The analysis first focuses on a two-node network where there are cheap generation supplies available in an exporting region, expensive generation supplies controlled by a single firm in an importing region, and a congested transmission link between the two regions. We find that holding financial rights enhances the market power of a monopoly generator in the importing region. The ultimate allocation of rights depends on the microstructure of the rights market through the ability of initial rights holders to free ride on generators with market power enhancing the value of the transmission rights. We next examine whether and how reliance on physical rather than financial rights affects these results. We find that holding physical rights can both enhance the market power of a generator in the importing region and lead it inefficiently to restrict imports of cheap power from the exporting region by "withholding" some physical rights from the rights market. Inefficient withholding of physical rights leads us to consider "capacity release" rules which mitigate withholding. These results are summarized with a comparison of the welfare properties of financial and physical rights with and without capacity release rules. The paper concludes with discussions of extensions to alternative buyer and seller market power configurations and to a three-node network to incorporate loop flow considerations. The extension to the three-node network does not change the basic conceptual results, but does reveal complications with the use of physical rights in the presence of loop flow.

JEL Classification: L13, L94
Keywords: Power markets, electricity, transmission, congestion, transmission rights

*The authors are grateful to Kyle Bagwell, Bill Hogan, Steve Stoft, Bob Wilson and three referees for very helpful comments. This paper is a revised and combined version of two earlier papers entitled "Transmission Rights and Market Power on Electric Power Networks I: Financial Rights" and "Transmission Rights and Market Power on Electric Power Networks. II: Physical Rights".

†Department of Economics, and Center for Energy and Environmental Policy Research, MIT.
‡IDEI and GREMAQ (UMR 5603 CNRS), Toulouse, CERAS (URA 2036 CNRS), Paris, and MIT.

1 Introduction

There has been considerable controversy over whether competitive electricity systems should be organized around bid-based pools with financial transmission rights or bilateral contracting systems organized with tradeable physical transmission rights (Joskow (1996)). We focus here on one set of issues that have arisen in this controversy. We analyze whether and how the allocation of transmission rights associated with the use of an electric power network affects the behavior of electricity generators and purchasers which possess market power. We examine the similarities and differences in this regard between financial and physical rights and compare the welfare properties of each. We also examine how transmission rights markets with different microstructures allocate rights among generators and consumers and determine rights prices, and we demonstrate that the allocation of rights through the market is endogenous.[1]

Electricity supply has traditionally been characterized by vertically integrated monopolies subject to public regulation. That is, the generation, transmission and distribution of electricity in a particular geographic area has typically been the responsibility of a single regulated firm. However, many countries recently have restructured or are in the process of restructuring their electricity sectors. One of the primary goals of electricity sector restructuring is to create unregulated competitive generation services markets with many competing generation suppliers and open entry. The transmission network, however, typically remains a regulated monopoly.

Competing generators must rely on the transmission network to schedule and dispatch their plants to support sales of electricity in organized spot and forward markets and through bilateral contracts with end-use customers or marketing intermediaries (including distribution companies) which in turn supply end-use customers with electricity. Since a transmission

[1]Our analysis is limited to these issues and it is not our objective to discuss here the full set of reasons why a physical rights mechanism might be preferred to a financial rights mechanism or vice versa. The relevant considerations that must be taken into account in evaluating the alternative organizational models include the transactions costs associated with using, trading and enforcing the different types of rights to facilitate efficient supply of electricity service given network constraints, and the need to adjust the supply of rights to reflect rapid changes in the quantity of network capacity actually available at any particular point in time. The analysis of this broader set of issues is beyond the scope of this paper.

operator which also has a financial interest in the generation sector may have incentives to discriminate against competing generators, the generation of electricity is typically separated from the operation of the transmission network. We will refer to the transmission network operator created through such a restructuring process as the Independent System Operator (ISO). In most countries that have restructured, the ISO also owns the transmission assets, and is typically referred to as a "Transco." In portions of the U.S., however, separate non-profit ISOs have been created to operate transmission assets owned by other firms.

The demand for electricity varies widely from hour to hour, day to day and month to month. In addition, transmission facilities are sometimes out of service due to equipment failures or scheduled maintenance. As a result, during some time periods, one or more lines on the network can become congested and cannot accommodate all of the power flows that would occur if the transmission capacity constraints did not exist. As a result, mechanisms must be developed to allocate scarce transmission capacity in an efficient manner. In addition, transmission congestion has potential implications for the intensity of competition in generation markets. For example, when demand is very high in the summer in San Diego, CA there typically does not exist enough transmission capacity to fully satisfy this demand with imports of relatively low-cost supplies from generators located elsewhere in the Western United States. San Diego becomes what is sometimes called a "load pocket" at these times. A small number of older relatively inefficient generators located in San Diego must be relied upon to balance supply and demand. Because there are few generators inside the congested area, these generators may have market power when imports are constrained.[2]

There are two general approaches to allocating scarce transmission capacity. One popular approach that is increasingly being used in the U.S. and other countries is built on an industry model in which the network operator manages organized public forward and spot electricity markets into which generators can submit minimum bids to supply and buyers maximum bids

[2] "Market Power in the San Diego Basin," California ISO, December 1999. Contractual mechanisms have been applied to these generators in an effort to mitigate their market power but it has proven difficult to design contracts that do so efficiently. More generally, seller market power issues have been the focus of considerable attention in the electricity sectors in the U.S. and the UK. See for example Wolfram (1999) and Bushnell and Wolak (1999). In California, generators have expressed concerns about the buyer market power of distribution companies that purchase energy in the day-ahead market.

to purchase electricity at specific locations or "nodes" on the network. The network operator then chooses the lowest cost bids to balance electricity supply and demand subject to the physical laws that govern electric power networks and the capacity of the network to carry power reliably. The bid price of the last bidder selected (or the first bidder rejected) at a node becomes the market clearing price at this node. When the transmission network is congested, market clearing prices will vary among locations or nodes on the network. Prices are higher at locations that are import constrained and lower at locations that are export constrained. The differences between locational prices represent congestion charges that generators at low-priced locations pay to supply power to customers at high-priced locations. Since demand and transmission capacity availability both vary over time, the incidence of network congestion,the differences in locational prices, and congestion charges can also vary widely over time. The associated variations in prices create a demand by risk averse buyers and sellers for instruments to hedge price fluctuations to provide them with a "firm price" for transmission service.[3] In order to satisfy this demand, several ISOs in the U.S. have created and allocated "financial rights" to market participants. These financial rights give the holders a claim on the congestion rents created when the network is constrained and allow them effectively to hedge variations in differences in nodal prices and associated congestion charges.

A second approach to allocating efficiently congested network interfaces is to decentralize congestion pricing by creating and allocating another type of tradable transmission rights which give a holder *physical* rights to use congested transmission interfaces. Under this approach, the physical capacity of each of the potentially congested interfaces is defined and rights to use this capacity are created and allocated in some way to suppliers and consumers. A supplier must possess a physical right to have its supplies scheduled or "transported" over the congested interface. Once it has such a physical right, there is no additional charge for using the congested interface. The markets for these physical rights then determine the market clearing prices for congestion. Holding physical transmission rights also plays the same role as do financial rights

[3] A form of risk aversion is generated by firms' need to insure against liquidity shocks when financial markets are plagued by agency problems and credit rationing. See Holmström-Tirole (1998). In the U. S., electricity marketers and unregulated generators have aggressively promoted the need for "firm transmission rights" that will allow them to insure themselves against the costs of network congestion.

in hedging variations in congestion prices since rights holders pay no additional congestion charges. In this organization model, the ISO's role is much more passive, relying primarily on bilateral contracting and private auction markets to determine which generators are dispatched at various times and how scarce transmission capacity is allocated.

Our analysis focuses first on a two-node network where there are cheap generation supplies available in an exporting region (the "North"), expensive generation supplies in an importing region (the "South"), and a congested transmission link between the two regions. We recognize that this is a highly simplified characterization of an electric power network. However, it is important to understand the effects of different types of transmission rights in a simple network before we go into more complex networks. Moreover, the two-node network model captures important congestion attributes of real electric power networks or sub-networks in England and Wales, Argentina, Chile and New Zealand. It is also in this type of simple network where physical rights are the most likely to represent a practical alternative to financial transmission rights because the transactions costs and network externality problems associated with physical rights are less severe in such a simple network.

We find that *if* the generator(s) in the importing region have market power, their holding financial rights enhances that market power. However, the analysis recognizes that the ultimate allocation of rights to market participants is endogenous and depends on the microstructure of the rights market. Three alternative market microstructures are examined. They vary in the extent to which initial rights holders can free ride on the ability of generators with market power to use it to enhance the value of the transmission rights. Except when there is full free riding, a generator with market power in the importing region is likely to acquire at least some financial rights.

We next examine whether and how reliance on physical rather than financial transmission rights affects these results. We find that holding physical rights can both enhance the market power of a generator in the importing region and lead it inefficiently to restrict imports of cheap power from the exporting region by "withholding" some physical rights from the rights market. The extent of withholdings depends on the ability of the generator with market power to also act as an intermediary delivering imports from the North to consumers in the South. Inefficient withholding of physical rights leads us to consider "capacity release" rules to mitigate withholding and other regulatory indicia to identify market power enhancing effects of both financial and physical rights. Our results for the effects of financial and physical rights for this market power configuration are summarized with a comparison of the welfare properties of financial and physical rights with alternative assumptions about capacity release rules and commitment. A striking conclusion of this analysis is that from the perspective of the effects of transmission rights on market power and production efficiency, the absence of transmission rights created by the ISO does as well or better than either type of transmission rights system.

The paper concludes with a summary of two extensions. The first examines alternative buyer and seller market power configurations. Transmission rights holdings may not affect generator market power in the exporting region, decrease buyer market power in the importing region, and increase buyer market power in the exporting region. We then examine a three-node network to incorporate loop flow considerations.[4] The extension to the three-node network enables us to analyze the externalities created by loop flow; such externalities are substantial for instance in Continental Europe[5] and some of the United States. It does not change the basic results, but does reveal interesting interactions between generators at different nodes as well as complications with the use of physical rights in the presence of loop flow.

[4]The two-node network is used to capture essential features of "radial networks" in which dispersed generators supply energy directly to consumption nodes without direct physical interactions between them. The three-node network is used to capture essential features of "loop flow" on electric power networks. In reality, electric power networks often have even more complicated combinations of radial and loop flow features and power flows can change directions at different points in time. However, the two-node and three-node cases are the best places to start to understand the core attributes of market power and the interaction with transmission rights on electric power networks.

[5]See the Haubrich-Fritz report for the European Commission (DG17, 1999).

2 A simple electricity model with congestion

2.1 The two-node network under perfect competition

2.1.1 Bid-based dispatch and financial rights

Let us consider first a restructured electricity sector that consists of a group of unintegrated and unregulated generating companies and an ISO which operates the transmission network, manages a spot energy market, and dispatches generators based upon their bids to supply generation services so as to balance the supply and demand for generation services in an efficient manner, taking into account physical constraints on the transmission network. The demand side can be thought of either as demand placed in the wholesale market by distribution companies which then resell the generation services to end-use consumers or as retailing intermediaries who buy energy in the wholesale market and then resell directly to end-use consumers who have paid for access to "unbundled" distribution "wires" services. Like in all existing bid-based dispatch systems, actors bid supply (or demand) schedules and receive (or pay) the same location-specific price for all units supplied (or purchased). Furthermore, we can assume, without loss of generality in our model, that a monopoly or monopsony at a node sets the price at that node, since it can always manipulate that price by selecting an appropriate supply or demand schedule.

If the suppliers and buyers behave competitively, the bidding process truthfully reveals the marginal cost curves and demand functions to the ISO. The ISO knows the physical constraints on the transmission network and, therefore, is in a position to enable an efficient dispatch of the generators given transmission constraints. The efficient dispatch may exhaust the capacity on some links, leading to congestion and congestion charges for using the congested link defined by the difference in nodal prices for electricity. This industry structure goes along with a financial transmission rights system which players can rely on to hedge the uncertain costs associated with congestion charges.

Sections 2 through 6 consider a simple two-node (no loop flow) network,depicted in Figure 1, where there are a set of low-cost generators (G_1) in the North which produce output q_1

and have an aggregate cost function $C_1(q_1)$, with $C_1' > 0$ and $C_1'' > 0$.[6] We assume in these sections that these generators behave competitively when they submit supply bids to the ISO. There is no demand in the North and we refer to the North as being either the upstream location or the exporting region. The market clearing price for generation sold in the North is p_1. In the South, there are electricity consumers and a set of generators (G_2) that have higher production costs (within the relevant range) than do the generators in the North. We refer to the South as the downstream location or the importing region. We assume in this section only that these generators too behave competitively. The market clearing price for generation in the South is p_2. Consumers have a demand function $Q = q_1 + q_2 = D(p_2)$ with $D' < 0$ and where p_2 is the price for all generation service paid by consumers in the South.

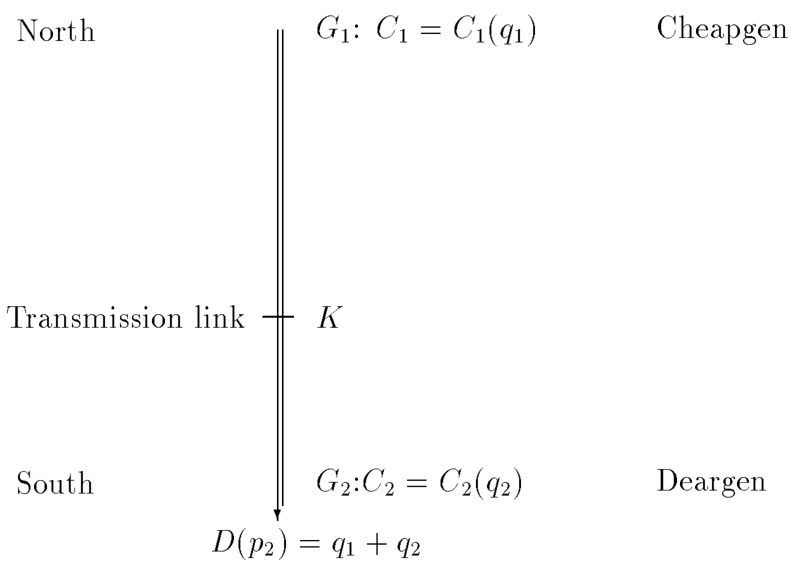


Figure 1

Finally, there is a transmission line linking the North with the South which has a fixed capacity equal to K. The nondepreciated capital and operating costs of this link are assumed to be recovered separately from consumers in lump sum charges net of revenues produced by

[6] Short run supply (marginal cost) functions for electricity are generally upward sloping, reflecting diversity in the efficiency with which individual generating units transform primary fuels into electricity. However, short run supply functions may have long segments that are fairly flat reflecting multiple generating units with similar transformation efficiencies. Accordingly, marginal cost may be locally flat or upward sloping depending on the level of demand and the attributes of the generating units available to supply electricity at a particular point in time.

selling physical or financial transmission rights[7] and we do not consider these costs further in our analysis. To further simplify things, we also ignore thermal losses on the network. We focus on situations where demand is sufficiently high that it cannot all be fully served by generators in the North because the transmission capacity constraint is binding. That is, some supplies from the less efficient generators in the South are required to balance supply and demand at the competitive prices. Thus, the marginal cost of generation in the North must be lower than the marginal cost of generation in the South when K is binding and "nodal prices" p_1 and p_2, which under perfect competition will be equal to the marginal costs in the North (C_1') and the South (C_2') respectively, will differ from one another with $p_1 < p_2$. Note, however, that consumers in the South pay the nodal price p_2 for all of their consumption $(Q = q_1 + q_2)$. Accordingly, scarcity rents $K(p_2 - p_1)$ accrue to the ISO, absent an alternative allocation of rights to collect these rents. We refer to these rents below as the ISO's "merchandising surplus."

Under these assumptions the nodal prices in the North and the South are given by the following equations:

$$
\begin{aligned}
p_1^* &= C_1'(K), \\
p_2^* &= C_2'(D(p_2^*) - K)) > p_1^* \\
q_1 &= K \\
D(p_2^*) &= q_1 + q_2 = K + q_2.
\end{aligned}
$$

In this case, the ISO (effectively) sells $K + q_2$ units of output for p_2^* per unit to consumers, but only pays out $p_1^* q_1 + p_2^* q_2$ to the generators. Accordingly, it earns a "merchandising surplus" of $(p_2^* - p_1^*)K$. One can alternatively think of these net revenues as congestion payments $(p_2^* - p_1^*)$

[7] In both California and in the Pennsylvania-New Jersey-Maryland (PJM) restructured systems, revenues from sales of transmission rights are returned to the owners of the transmission capacity to help to defray their capital and operating costs. The ISO, not the transmission owner defines the quantity of rights to be auctioned for each potentially congested link and defines the auction rules. The revenues are then treated as pure pass-through credits to the transmission charges paid by end-use consumers connected to the network.

made by the suppliers in the North to use the scarce transmission interface.[8]

Now we define more precisely what we mean by "financial" transmission rights:

Financial rights : Financial rights give the holders a proportionate share of the congestion payments or merchandising surplus received by the ISO when the transmission constraint K is binding. The owners of these rights may be generators, consumers, or speculators. Generators do not require financial rights to be dispatched by the ISO. But without them they must pay congestion charges when they supply over a congested link. There are K rights issued and the owner of one unit of financial rights entitles the owner to $\eta = (p_2 - p_1)$, or the difference in the nodal prices. Total payments given to rights holders are equal to $(p_2 - p_1)K$ ($(p_2^* - p_1^*)K$ in equilibrium).

If there is no market power in the generation market, the introduction of financial rights into this simple system has no effect on the prices for energy or the allocation of resources. The ISO's revenue from congestion rents or its merchandising surplus is now transferred to the owners of financial rights. The nodal prices for energy are as defined above and the competitive market value of the financial transmission rights is simply equal to the difference in nodal prices. Nodal energy prices and the price of financial rights are therefore given by

$$p_1^* = C_1'(K)$$

$$p_2^* = C_2'(D(p_2^*) - K) > p_1^*$$

$$\eta^* = p_2^* - p_1^*.$$

These equilibrium conditions for a competitive electricity market with tradable financial transmission rights provide a benchmark against which we can compare other market outcomes. In particular, we want to explore the interaction between market power in the electricity market

[8] In what follows we ignore price uncertainty and focus on the effects of market power on the allocation of transmission rights and the associated equilibrium prices, quantities, and production costs with different market microstructures. Most of the previous literature on which we build (Bushnell 1998, Oren 1997, Hogan 1992) also ignore uncertainty and simply focus on situations when the link(s) is congested. This may seem odd, however, since one motivation for both financial and physical rights is to create instruments that allow buyers and sellers to "hedge" variations in profits (due both to variations in equilibrium supply prices and/or congestion charges) caused by congestion on the network. It is useful to focus first on how market power affects the demand for rights and, given the characterization of market power, how rights markets with different microstructures allocate these rights.

and the allocation of financial transmission rights to sellers and buyers of electricity and how alternative microstructures for the rights market ultimately affect the allocation of rights among electricity sellers and buyers under alternative assumptions about market power.

2.1.2 Bilateral contracts and physical rights

Let us now turn to the case of bilateral contracts and physical rights. Generators in the North must have physical rights to schedule their generation pursuant to their bilateral contracts with consumers in the South. Since the transmission capacity is a binding constraint, rights to use it have a market value (η) that is greater than zero. The net price p_1 (net of the cost of physical rights) generators in the North receive for their generation supplies is then simply the difference between the price they are paid by their customers in the South (p_2) minus the market value of the physical transmission rights they need to deliver it (η). (Alternatively, we can think of p_1 as the price of generation produced and delivered in the North. Retail marketing intermediaries would acquire the generation in the North, acquire the necessary transmission rights in the rights market at a market price η, and schedule the supplies with the ISO for delivery to their customers in the South who pay a delivered price p_2.) Generators in the South do not need physical transmission rights since they do not use the transmission line as a result of their proximity to consumers and receive both a gross and net price equal to the delivered price in the South p_2.

When the energy and rights markets are perfectly competitive the equilibrium conditions are as follows:

$$p_2^* = C_2'(D(p_2^*) - K),$$

$$p_1^* = C_1'(K),$$

$$p_2^* > p_1^*,$$

$$p_1^* = p_2^* - \eta^* \iff \eta^* = p_2^* - p_1^*,$$

$$D(p_2^*) = q_1 + q_2 = K + q_2.$$

These are the same equilibrium conditions that emerged under perfect competition with a financial rights system. The price of physical and financial rights are the same, the delivered price in the South is the same and the price received by generators in the North net of the cost of physical rights is equivalent to the nodal price in the North derived under bid-based dispatch and nodal pricing in our companion paper. This verifies for our model the more general result (due to Chao-Peck 1996) concerning the equivalence of financial and physical rights when the energy and rights markets are perfectly competitive.

2.2 Market power at the expensive node: the no-rights benchmark

Most of our analysis examines the case where the generators in the South (G_2) are owned by a single firm that has market power, while the generators in the North (G_1) behave competitively. This is a typical case which arises in many urban areas.[9] The single generator in the South now maximizes profits given its residual demand curve, defined by consumer demand for electricity in the South net of the supply of electricity from the North. The price in the South is higher than in the perfectly competitive environment and so the transmission link is congested ($q_1 = K$) a fortiori.

The nodal price, quantities produced, and generator profits in the North are the same as in the competitive cases analyzed earlier:

$$p_1 = C_1'(K) = p_1^*.$$

However, G_2 now chooses p_2 by maximizing profit against its *residual demand curve*

$$q_2 = D(p_2) - K.$$

G_2's profit function is the profit associated with generation supplies:

$$\mathcal{G}(p_2) = p_2[D(p_2) - K] - C_2(D(P_2) - K).$$

We assume that $\mathcal{G}(\cdot)$ is strictly concave. The profit maximizing price p_2^m is higher and the quantity q_2^m produced in the South lower than they would be if G_2 behaved competitively:

$$p_2^m > p_2^* \text{ and } q_2^m < q_2^*.$$

[9] Generators initially were often developed close to city-centers. As generators became larger, urban sites became scarce, and transmission technology improved, newer more efficient generators required to meet growing demand tended to be sited further from load centers and are more widely dispersed geographically.

Now, assume that either financial rights or physical rights have been defined and allocated. If G_2 maximizes profits without taking any impact on the value of rights into account, the competitive market price of both the financial and physical rights is given by

$$\eta^m = p_2^m - p_1^*.$$

3 Financial rights and market power

3.1 Impact of financial rights ownership on market power

We continue to assume that there is a single generator in the South with market power but allow this generator to hold financial rights. Does holding these rights increase G_2's market power? Recall that the value of financial rights is given by the difference between the nodal prices for energy in the North and in the South. Since market power in the South increases energy prices in the South, congestion rents and the value of financial rights,

$$\mathcal{F}(p_2) = (p_2 - p_1^*)K,$$

must increase as well. That is, when the transmission link is congested, the value of financial transmission rights varies directly with the contraction of output and the increase in delivered energy prices in the South associated with the exercise of market power in the South.

Assume that G_2 holds a fraction $\alpha_2 \in [0, 1]$ of the K financial rights available. It now faces the following profit function to maximize:

$$\begin{aligned}
\pi_2(\alpha_2) &= \max_{p_2} \{\mathcal{G}(p_2) + \alpha_2 \mathcal{F}(p_2)\} \\
&= \max_{p_2} \{p_2 [D(p_2) - K] - C_2 (D(p_2) - K) + \alpha_2 [p_2 - C_1'(K)] K\} \\
&= \max_{p_2} \{\Pi_2 (\alpha_2, p_2)\}.
\end{aligned}$$

Note that $\partial \Pi_2 (\alpha_2, p_2) / \partial \alpha_2 \partial p_2 = K > 0$: The larger the fraction α_2 of rights held by the generator, the stronger its incentive to jack up the price in the South. The optimum $p_2(\alpha_2)$ is increasing continuously[10] in α_2 from $p_2(0) = p_2^m$ to $p_2(1)$ (which maximizes $\{p_2 D(p_2) -$

[10] $\mathcal{G}(\cdot) + \alpha_2 \mathcal{F}(\cdot)$ is strictly concave since we assumed that $\mathcal{G}(\cdot)$ is.

$KC_1'(K) - C_2(D(p_2) - K)\}$. The profit maximizing price for G_2 to set for energy in the South increases directly with α_2, and as α_2 increases the quantity produced by G_2 decreases as well. G_2 now has two revenue streams: one stream of revenue from sales of energy ($\mathcal{G}$) and a second stream of revenues from the congestion rents ($\mathcal{F}$) that it is entitled to by virtue of holding the financial rights. The more G_2 internalizes the congestion rent, the higher the congestion rent by virtue of G_2's control of p_2. When G_2 has financial rights, it effectively reduces the elasticity of the residual demand curve and increases its market power. Let

$$\eta(\alpha_2) \equiv p_2(\alpha_2) - C_1'(K) = p_2(\alpha_2) - p_1^*.$$

When $\alpha_2 = 1$, the monopoly in the South faces the *total demand* ($D(p_2)$) rather than the residual demand ($D(p_2) - K$) it faces when it holds no financial rights. That is, if the monopoly generator in the South holds all the financial rights, it maximizes its profit (G_2's net revenues from supplying energy *plus* its revenues from congestion rents) as if it had a monopoly over the entire demand function. In doing so, G_2 sacrifices some profits it would otherwise earn from supplying electricity ($\mathcal{G}$) in order to increase the profits it receives in the form of "dividends" ($\mathcal{F}$) on the financial rights it owns as a result of its ability to increase the price p_2 in the South.

3.2 Microstructure of markets for financial rights

We have shown that if G_2 holds financial rights, these rights will enhance its market power. And the larger is the fraction of the rights it holds the greater is its market power. However, G_2 must acquire these rights through some type of market allocation mechanism. Accordingly, we want to explore the allocation and pricing of these rights when G_2 has market power and how they are affected by the microstructure of the rights market. In what follows, we assume that consumers in the South are not in the market for financial rights, or, if they are, that the ownership of these rights is too dispersed to create countervailing power to G_2 in the purchase and sale of the rights.

From our discussion of the effects of ownership of financial rights on G_2's behavior we know that the value of financial rights increases with G_2's holdings of such rights. In this respect, G_2

resembles the raider or large shareholder of the corporate finance literature (Grossman-Hart 1980, Shleifer-Vishny 1986, Admati et al 1994, Burkart et al 1998).[11] G_2 would like to get all of the financial rights, but pay as little as it can for them. We know from the corporate finance literature that the realization of gains from trade between the initial holders of the financial rights (shares) and the value enhancing raider (G_2) depends on the extent of free riding and therefore on the microstructure of the rights market. Here, the initial financial rights (share) holders would like to hold on to their rights and capture their full value resulting from a larger difference in nodal prices. However, the value of the rights is maximized only if G_2 acquires all of them. As we shall see, the more initial holders of rights can free ride on the ability of G_2 to increase the value of these rights by exercising market power, the fewer rights is G_2 likely to acquire through the rights market.

There is a wide variety of possible trading structures for both financial rights and physical rights. We consider three examples of rights market microstructures here and in the following section on physical rights which (a) are interesting in their own right and (b) represent two polar cases and one intermediate case of free riding by initial rights holders. The three cases are:

a) *No Free Riding*: Rights are initially held by a *single owner* who is neither a generator nor a consumer. The single owner bargains with potential purchasers over the price at which the rights will be transferred.

b) *Full Free Riding*: In this case the initial ownership of rights is dispersed among non-stakeholders and stakeholders without market power. G_2 makes a *tender offer* at some price η. The tender offer is unconditional.

c) *Partial Free Riding*: Here we assume that all of the rights are auctioned off to the highest bidders by the ISO, as proposed for California.[12]

[11] A slight difference (which simplifies the analysis) with the raider in a corporate control situation is that the impact of holdings moves continuously instead of jumping at 51% of the rights.

[12] The finance literature has looked at alternative environments in which the gains from trade between the raider and the initial shareholders are partially realized. In Kyle-Vila (1991) the presence of liquidity traders allows the raider to disguise her purchases (as long as she restricts her order flow) and prevents full free riding. In Holmström-Nalebuff (1992), the value of the firm increases discontinuously when the raider obtains a controlling

We also want to examine the effects of allowing consumers (distributors acting as agents for their end-use customers or end-use customers directly) to buy financial rights. We discuss each of these cases in turn.

3.2.1 Financial rights initially held by a single nonstakeholder owner (no free riding)

In this case, since G_2 has the highest value for the rights (in terms of profits, not total surplus) in the absence of bidding by consumers, G_2 will acquire all of the rights at a price negotiated with the initial owner. Accordingly, G_2's market power is enhanced and the equilibrium price p_2 is higher than it would be in the absence of financial rights. The negotiated market price for the financial rights will lie somewhere between $\eta^m = \eta(0)$ and $\eta(1)$. G_2 and the initial owner share the "surplus"

$$[\mathcal{G}(p_2(1)) + \mathcal{F}(p_2(1))] - [\mathcal{G}(p_2(0) + \mathcal{F}(p_2(0))].$$

The division of the monopoly rents associated with G_2's ownership of all of the rights depends on the relative bargaining power of G_2 and the initial owner.

Remark 1: It is important that the financial rights be in *positive net supply* through the market. Suppose in contrast that the congestion rents are allocated to a party who is prevented from selling them and participating in the financial market (i.e. the owner cannot contract directly or indirectly with G_2). The rights would then de facto be in zero net supply. Would a group of investors want to enter into a "gambling contract" with G_2 specifying that the investors will pay $(p_2 - p_1)t$ to G_2 once nodal prices are realized, where $t > 0$ is the scale of the financial deal ("gambling" refers to the fact that G_2's two components of profit $\mathcal{G}(p_2)$ and $(p_2 - p_1)t$ both increase with p_2)? The answer is no: The increase in the aggregate profit of G_2 and the investors, $\mathcal{G}(p_2(\alpha_2)) - \mathcal{G}(p_2(0))$, where $t = \alpha_2 K$ is the size of the side deal, would then be negative.

Remark 2: We show in Joskow and Tirole (1999a) that if consumers in the South can solve their

share. It is shown that if initial shareholders hold several shares each, then in a symmetric (mixed-strategy) equilibrium of the tendering subgame, shareholders cannot fully free ride.

collective action problem and form a coalition, they outbid G_2 for the rights. This is the case because there is dead-weight loss to consumers in the South as a result of the enhancement of G_2's market power and consumers' willingness to pay exceeds the increased profits that would accrue to G_2. Note however, that the consumer coalition appears to "lose money" by buying the rights since the amount the coalition pays for the rights would be greater than their value ex post. If the agent for the consumers is a regulated distribution company it could face regulatory penalties if in a prudency review the regulator naively compared only the price paid for the rights with their ex post value and ignored the value of market power mitigation.

3.2.2 Tender offer by G_2 (full free riding)

It is clear that when G_2 makes an unconditional tender offer to dispersed owners (without market power in either market) who initially hold the rights, it does not want to purchase any rights: Suppose G_2 offers to buy whatever is tendered at a price η such that

$$\eta(0) < \eta \leq \eta(1).$$

The fraction α_2 of rights tendered is given by:

$$\eta = \eta(\alpha_2) = p_2(\alpha_2) - p_1^*.$$

G_2's profit is then given by

$$\mathcal{G}(p_2(\alpha_2)) + \alpha_2\mathcal{F}(p_2(\alpha_2)) - \alpha_2\eta K = \mathcal{G}(p_2(\alpha_2)) < \mathcal{G}(p_2(0)),$$

where $p_2(0)$ maximizes G_2's profits from supplying generation service.[13]

The point made here is completely general. Any player who makes a tender offer buys the rights at a price equal to their ex post price. The value of the rights goes up or at least stays constant after a stakeholder with market power (consumer or producer in the North or

[13]With conditional offers, we are back to the absence of free riding: G_2 can offer to pay $\eta = p_2(0) - p_1^* + \varepsilon$ (where ε is small) and stipulate that the offer is valid only if *all* rights are tendered.Then everyone tenders. Note, though, that with a conditional offer it makes a difference whether the rights are held by financiers or by stakeholders (we are grateful to Bruno Biais for this point). The no-free riding point holds only if the rights are held by financiers. Suppose for example that a competitive consumer in the South holds a single right. By refusing to tender this right the consumer forgoes the profit ε on the sale, but lowers the price in the South from $p_2(1)$ to $p_2(0)$ by defeating the tender offer.

the South) purchases a fraction of the rights from nonstakeholders or stakeholders with no market power. However, the utility of the stakeholder with market power purchasing the rights decreases and, as a result, it will purchase no rights in this case. For example, G_2 must sacrifice some profits associated with the supply of generation service in order to jack up the price of energy p_2. However, it cannot recoup these lost profits through the higher dividends on any rights it buys resulting from such a price increase because the value of these dividends would be reflected in the price it would pay for the rights. Since there is full free riding on any enhanced value of rights that G_2 can create by further contracting its output and increasing p_2, it is not profitable for it to buy any rights.

3.2.3 Auctioning of the rights by the ISO (partial free riding)

Assume that there is no consumer coalition and the ISO auctions off all of the rights simultaneously. We analyze a *discriminatory* auction, that is an auction in which i) bidders announce a price and a maximum quantity they are willing to buy at this price, ii) rights are allocated to the highest bidders[14], and iii) bidders pay their bids. We assume that the market is deep in the sense that risk neutral arbitrageurs, the market makers, stand ready to arbitrage away any profit opportunity.

Note first that G_2's bid cannot be deterministic. Suppose G_2 bids $\eta > \eta(0)$ and purchases $\alpha_2 K$ rights. Either $\alpha_2 = 0$ and then market makers overpay for the rights (they pay above η for rights whose value is $\eta(0)$). Or $\alpha_2 > 0$, and $\eta \geq \eta(\alpha_2)$ (if $\eta < \eta(\alpha_2)$, then a market maker could make a profit by bidding for one right at a price between η and $\eta(\alpha_2)$); so G_2's profit is at most $\mathcal{G}(p_2(\alpha_2)) < \mathcal{G}(p_2(0))$. That is, G_2 would be better off not bidding for rights. But if G_2 does not bid for rights, market makers (or stakeholders without market power) are willing to pay $\eta(0)$, in which case G_2 can buy all rights at a price just above $\eta(0)$ and increase its profit, a contradiction. Hence G_2 must randomize in equilibrium. Furthermore, G_2 optimally buys all rights available at his bid.[15]

The market makers face a *winner's curse* problem: They tend to get rights precisely when

[14] The rationing rule used in case of a tie will turn out to be irrelevant.

[15] This is a consequence of the fact that his profit is convex in the number of rights purchased at a given price.

G_2 does not and so when rights are not very valuable. The consequence of this winner's curse is that *the competitive market makers' demand function* is not flat at some price, but rather *downward sloping*. A higher bid by a market maker is costly and so must be compensated by a higher value of the right conditionally on the bid being a winning bid. The distribution of the number of rights held by G_2 conditionally on bid η being a winning bid indeed shifts to the right when η grows, if the market makers' demand function is downward sloping.

In equilibrium G_2 randomizes over the interval $[\eta(0), \bar{\eta}]$, where $\eta(0) < \bar{\eta} < \eta(1)$, according to density $h(\eta)$ and cumulative distribution function $H(\eta)$. [That is, the probability that G_2's bid is less than η is $H(\eta)$.] The market makers' aggregate demand is given by a decreasing function $\hat{d}(\eta)$, with $\hat{d}(\eta(0)) = K$ and $\hat{d}(\bar{\eta}) = 0$. For the purpose of the analysis, it will be convenient to define the fraction of rights $\hat{\alpha}_2(\eta)$ that is acquired by G_2 when bidding η:

$$\hat{\alpha}_2(\eta)K = K - \hat{d}(\eta),$$

with $\hat{\alpha}_2' > 0$, $\hat{\alpha}_2(\eta(0)) = 0$, $\hat{\alpha}_2(\bar{\eta}) = 1$.

G_2's behavior in the rights market

In order for G_2 to be indifferent between all bids in $[\eta(0), \bar{\eta}]$, it must be the case that they all yield G_2 the same profit, equal to the profit $\mathcal{G}(p_2(0))$ obtained by bidding $\eta(0)$ and obtaining no rights:

$$\mathcal{G}(p_2(0)) = \mathcal{G}(p_2(\hat{\alpha}_2(\eta))) + \hat{\alpha}_2(\eta)\mathcal{F}(p_2(\hat{\alpha}_2(\eta))) - \eta\hat{\alpha}_2(\eta)K. \tag{1}$$

Equation (1) defines an increasing function $\hat{\alpha}_2(\eta)$ and thereby a decreasing demand $\hat{d}(\eta)$ by the market makers. The upper bound of the support of G_2's strategy, $\bar{\eta}$, is given by

$$(\eta(1) - \bar{\eta})K = \mathcal{G}(p_2(0)) - \mathcal{G}(p_2(1)).$$

Market makers' zero-profit condition

Consider a market maker playing a bid for one right at price $\eta \in [\eta(0), \bar{\eta}]$. With probability $1 - H(\eta)$, G_2's bid is higher and the market maker's bid is not selected. With probability $H(\eta)$, the market maker receives the right. His profit when G_2's bid is $\tilde{\eta} < \eta$ is $[p_2(\hat{\alpha}_2(\tilde{\eta})) - p_1^*] - \eta$, and so the zero-profit condition for all η can be written as: For all $\eta \in [\eta(0), \bar{\eta}]$,

$$\int_{\eta(0)}^{\eta} [[p_2(\hat{\alpha}_2(\tilde{\eta})) - p_1^*] - \eta]\, h(\tilde{\eta}) d(\tilde{\eta}) = 0.$$

This condition is obviously satisfied at $\eta = \eta(0)$. For it to be satisfied over the whole interval, the derivative of the left-hand side must be equal to zero, or

$$\frac{h(\eta)}{H(\eta)} = \frac{1}{[p_2(\hat{\alpha}_2(\eta)) - p_1^*] - \eta}. \tag{2}$$

Knowing $\hat{\alpha}_2(\cdot)$ from (1), equation (2) defines the bidding strategy $H(\cdot)$ for G_2.[16]

The number of rights purchased by G_2 is random and intermediate between those purchased under full free riding (none) and no free riding (all).

4 Physical rights and market power

4.1 Basic difference between physical and financial rights

Why might one expect there to be *any* differences between the effects of a financial rights system vs. a physical rights[17] system on the market power of G_2, other things equal? A potentially important difference is that unlike the case with financial rights, reliance on physical rights makes it possible that some market participants may find it profitable to withhold rights from the market, leading to a reduction in the effective capacity of the transmission link. Since generators in the North (or their customers in the South) must have rights to use the transmission link, the rights that they acquire effectively defines the capacity of the link (up to K). If the market leads to an allocation where generators at the cheap node (the North) do not end up holding all of the rights (K) and cannot (or do not) use all of the capacity (K) available on the link, then the supply of "cheap" power from the North available to meet

[16]Integrating (2) yields $H(\eta) = \exp(-\int_{\eta}^{\bar{\eta}} \ell(\tilde{\eta})d\tilde{\eta})$, where $\ell(\cdot)$ is the right-hand side of (2). Using (1) and the first-order condition defining $p_2(\alpha_2)$, it can be seen that $\ell(\eta)$ is of the order $1/(\eta - \eta(0))$ in the neighborhood of $\eta(0)$ and so $H(\eta(0)) = 0$.

[17]We here consider only plain vanilla physical rights. A referee suggested to us that it would be worth analyzing the case of physical rights tied to negative financial rights, that is financial rights that oblige their owners to pay when the line is congested.The idea is that such physical rights would reduce their owners' incentives to withhold, since withholding increase the congestion. More generally, it would be worth looking at mixed (financial plus physical) rights systems.

demand in the South would be reduced. In this case, supply from the cheap generators in the North (q_1) will be restricted ($q_1 < K$) and more demand than is necessary will be satisfied with expensive power from the South. Thus, "withholding" of rights from generators in the North could result in production inefficiency since expensive power from the South is substituted for cheaper power in the North.[18]

Quite generally, G_2 can attempt to capture *three rents* corresponding to the three markets (two local electricity markets and the rights market). The first rent is the consumer net surplus in the South; regardless of the institutional arrangement, G_2 has local monopoly power in the South and so the same ability to extract consumer surplus. Indeed, the price in the South always exceeds the price that maximizes generation profit in the South when G_2 faces the residual demand curve when the link is congested ($p_2 \geq p_2^m$). Thus, the action is with respect to G_2's impact on the other two rents/markets: value of rights and inframarginal rents of the competitive generators in the North. Financial rights do not enable G_2 to reduce the power flow from North to South and thus to reduce the inframarginal rents of generators in the North. In contrast, under physical rights, G_2 can withhold transmission capacity and thereby capture some of the inframarginal rents in the North. On the other hand, physical rights receive no dividend, and so, in contrast with the case of financial rights G_2 cannot affect the value of associated dividends. Physical and financial rights therefore do not allow G_2 to impact the same rent. It is remarkable then that the two systems can be compared so readily.

[18]The motivation for and implications of physical rights withholding here are different from those in Bushnell's (1999) recent paper. Bushnell assumes that production at both nodes is equally efficient and that both exhibit constant returns to scale. The only role that the transmission link plays in his model is to mitigate the generator's local market power at one of the nodes. This assumption seems less realistic than those made here. Historically, major interregional transmission lines were built to bring electricity from areas where it is cheap to produce to areas where it is more expensive to produce. They were not built to mitigate local market power problems. In the future, in restructured electric power sectors, transmission lines are expected to be built for the same reasons, unless regulators do not have other instruments available to mitigate local market power problems. Also, the mechanism here is different from that in Bushnell (1999). Bushnell assumes that there is a lot of capacity K on the link and that it is competition from generators using this link that keeps G_2 from exercising market power. Here we have assumed that capacity K is limited even when there is no generation market power and does not prevent the exercise of market power by G_2. Withholding of physical rights is motivated by the desire to extract rents from G_1.

We note as well that production inefficiency of the type we identify here is not possible in Bushnell's model since he assumes that the marginal costs of the generators are the same at each node.

4.2 Withholding of physical rights

We will consider in detail here only the situation in which a non-stakeholder owner initially owns all of the physical rights, the case of no free riding. This case allows us to identify how the market for rights takes into account two special characteristics of physical rights. As we later note, the impacts of the two alternative rights market microstructures are identical to what we found for financial rights above.

It is optimal for G_2 to acquire the rights in order to avoid non-internalized externalities between the two players. In essence, G_2 then produces electricity in two ways: first, by selling rights to generators in the North or by purchasing power from them and then keeping the rights to dispatch the power produced in the North, and, second, by producing power in the South. G_2 obtains the maximum profit by importing power from the North and reselling this power together with its own power to the consumers in the South. In contrast, if G_2 first sells $q_1 \leq K$ rights to generators in the North and then chooses its own production q_2 in the South, G_2 does not internalize in the latter decision the change in value of the rights sold earlier in the rights market. Because G_2 then sells power in two stages, it tends to overproduce in the electricity market ex post, in the same way Coase's durable good monopolist floods the market after having previously sold. The standard solution to Coase's durable good problem is leasing or vertical integration, which here corresponds to G_2's purchasing power in the North and thus keeping an exclusive relationship with the consumers in the South.

We are thus led to consider two cases:

Commitment:	G_2 imports power $q_1 \leq K$ from the North and sells $q_1 + q_2$ to consumers in the South.
Noncommitment:	G_2 cannot resell power produced in the North (say, because competition policy prohibits it). It sells $q_1 \leq K$ rights to producers in the North, who contract with consumers; G_2 cannot commit to a level of production q_2 in the South when selling rights to generators in the North.

4.2.1 Commitment

As we discussed, G_2's preferred outcome is obtained when it imports q_1 units (at price $C_1'(q_1)$ each) from the North, or, equivalently, when G_2 simultaneously sets a price p_2 for power in the South and a price η for rights (which it acquired earlier from the non-stakeholder owner). These two prices determine (in the relevant range) a quantity $q_1 \in [0, K]$ flowing through the congested interface, with

$$p_2 - C_1'(q_1) = \eta.$$

G_2's profit is

$$\max_{\{p_2, q_1\}} \left\{ p_2 \left[D(p_2) - q_1 \right] - C_2 \left(D(p_2) - q_1 \right) + \left[p_2 - C_1'(q_1) \right] q_1 \right\}.$$

Note that G_2 is a "gatekeeper" for production in the North when it controls all of the physical rights. It is both a monopsonist and a monopolist. It sells its own power and then it "outsources" to G_1 as well. So, G_2 faces a "make or buy" decision:

$$\begin{aligned} &\text{either } q_1 = K \\ &\text{or } q_1 < K \quad \text{and} \quad C_2'(D(p_2) - q_1) = C_1'(q_1) + q_1 C_1''(q_1). \end{aligned}$$

The term on the left of the latter equality is the marginal cost of (internal) production in the South and the term on the right is the " virtual marginal cost" (external) production in the North or the "perceived marginal cost" of G_2. In this case, G_2 finds it optimal to substitute expensive supplies from the South for cheaper supplies from the North in order to extract some inframarginal rents from the cheap generators in the North. This leads to production inefficiency.

Accordingly, when $q_1 < K$, p_2 will be higher than in the case where there is no generator market power at either node both as a result of withholding rights and as a result of the contraction of output in the South given q_1.

4.2.2 Noncommitment

Now assume that G_2 cannot buy in the North and resell in the South but must sell rights to generators in the North and cannot otherwise commit on its own ex post production when selling these rights. One may have in mind that the rights market operates first and then the power market (day-ahead or hour-ahead) operates given the distribution of physical rights arrived at the first stage. But the two markets can be simultaneous as long as G_2 is not able to demonstrate a credible commitment its own level of production when selling rights to generators in the North.

Power market: In the electricity market, G_2 takes q_1 as given and sets $p_2 = \hat{p}_2(q_1)$, where

$$\hat{p}_2(q_1) \text{ maximizes } \quad p_2[D(p_2) - q_1] - C_2(D(p_2) - q_1).$$

Rights market: In the first stage G_2 sells $q_1 \le K$ rights so as to maximize:

$$\max_{q_1} \{\hat{p}_2(q_1)[D(\hat{p}_2(q_1)) - q_1)] - C_2(D(\hat{p}_2(q_1)) - q_1) + [\hat{p}_2(q_1) - C_1'(q_1)]q_1)\}$$

given the function $\hat{p}_2(q_1)$.

Using the envelope theorem, the derivative of the latter objective function is equal to:

$$C_2'(q_2) - [C_1'(q_1) + q_1 C_1''(q_1)] + [d\hat{p}_2(q_1)/dq_1]\, q_1.$$

The third term, which is nonpositive, does not appear in the equivalent condition for the commitment case. It equals the change in value of the physical rights as downstream prices change.

4.2.3 Comparison of the commitment and noncommitment cases

The easiest case to examine is where there is constant returns to scale in the South (C_2' is constant). In this case, G_2 withholds "weakly" more physical rights in the noncommitment case.[19] This is a standard conclusion of "Coasian dynamics". Yet we cannot conclude that the commitment case dominates from a welfare point of view. It dominates from a production efficiency point of view, in that production from the cheap node is weakly greater than it is in

[19] "Weakly" comes from the fact that there may be corner solutions at $q_1 = K$. If $q_1 < K$ in the noncommitment case, then "strictly" is correct.

the noncommitment case. However, it does not dominate from a "market power" or downstream pricing point of view. In the noncommitment case, in the energy market G_2 ignores the effects of its production on the value of rights and "floods the market" to maximize profit given the output in the North which is defined by the rights that G_2 has sold in the first stage.

Example: Assume constant returns to scale in the North and the South ($C_1'' = C_2'' = 0$, $C_1' < C_2'$) and linear demand [$D(p) = 1 - p$]. Our analysis shows that there is no withholding under commitment. If $C_2' - C_1' > K/2$, then there is no withholding under noncommitment either.

Thus, if $C_2' - C_1' > K/2$, then noncommitment dominates since there is no withholding in either case and the downstream price p_2 is lower under noncommitment. Under these conditions physical rights dominate financial rights since the physical rights do not provide an additional incentive to G_2 to contract output in the energy market and, here, like financial rights do not generate withholdings and production inefficiency.

Let us conclude this section with a brief discussion of the impact of free riding in the rights market (see Joskow-Tirole 1999b for the formal derivations). With full free riding, G_2 has no incentive to acquire physical rights to enhance its market power. Since G_2 buys no rights, it maximizes its profits on the residual demand curve and physical rights do not enhance the market power of the monopoly generator in the South. This is the same result that we obtained for financial rights with this microstructure. With partial free riding, G_2 either acquires no rights (when the cost differential between the two regions exceeds some threshold) or plays a mixed strategy and acquires a fraction of the rights that lies between zero and one.

5 REGULATORY ISSUES

5.1 Physical rights and capacity release rules

One of the primary differences between a financial transmission rights system and a physical transmission rights system arises as a result of withholding of physical rights from the market which leads to an artificial contraction of the capacity of the transmission system. The potential for transmission capacity withholding naturally leads to the question of whether regulatory rules can be crafted which restrict the ability of stakeholders to withhold physical rights from the market. The transportation of natural gas on the interstate natural gas pipeline system in the U.S. is governed by a physical rights system. Pipelines are required to offer to enter into transportation contracts with gas shippers and gas consumers that give them the physical right to transport gas from one point to another on their pipeline networks. These physical rights are tradable, subject to regulatory price caps. Rights holders who do not use their rights to support the transport of gas by a certain time period prior to any particular transportation date are required to "release" those unused rights for sale to other shippers and consumers in the gas transportation market.[20]

Let's consider how a capacity release program might be implemented for electricity. Several issues need to be addressed. First, at what time in the generation scheduling process are physical rights deemed to be "unused" and available for release for use by other generators? Second, when an unused right is used by another user what, if anything, is the initial owner of the right paid for its use? Third, how does the system respond to an ex post realization that some rights that were designated for use in the scheduling process, and not made available under the capacity release program, are found not to have been used either due to conscious overscheduling by generators or due to unanticipated plant outages or reductions in consumer demand served under bilateral requirements contracts?

Counteracting physical transmission capacity withholding behavior that is a component of G_2's strategy to exercise market power in the electricity market, requires that the unused capacity be released for sale to competing generators in sufficient time that they can use the capacity effectively. In a regime governed solely by bilateral contracts between generators and consumers and a requirement that generators submit balanced schedules to the ISO, the value of the physical rights to competitors and the effects of their release on market power could be heavily influenced by how far in advance of the formal scheduling periods the rights are

[20]See U.S. Energy Information Administration, *Natural Gas 1996: Issues and Trends*, DOE/EIA-0560(96), December 1996, Washington, D.C.; U.S. Federal Energy Regulatory Commission, Notice of Proposed Rulemaking, *Regulation of Short-term Natural Gas Transportation Services*, July 29, 1998.

released and made available to others. It is difficult however to conceive of a pure bilateral contract system with a release program because there is then no natural date at which the bilateral market closes and the leftover capacity is released to allow...further bilateral trades. A realistic release program therefore seems to require a sequence of a bilateral market *followed by* a centralized bid-based auction market in which the released transmission capacity is made available to support additional supplies from generators in the North selected in the auction. The auction produces a set of market clearing spot prices in the North and the South as well as an allocation of generation supplies. Thus, we will consider a two-stage timing in which the bilateral market closes, say, a day ahead, and is then followed by bid-based dispatch for the remaining capacity:

Stage 1: *Bilateral market.* Bilateral contracts between buyers and sellers can be negotiated at any point of time (five years ahead, a year ahead, a week ahead...) before the date, say a day ahead, at which the balanced trades together with the associated physical rights must be registered with the ISO. Let q_1 denote the amount of power injected in the North as an outcome of the bilateral market.

Stage 2: *Bid-based dispatch.* The unused transmission capacity, namely $K - q_1$, is released. An auction market, run as described in section 3, opens with transmission capacity $K - q_1$. That is, the stage-2 market is the standard auction market except that the transmission capacity is reduced to the leftover capacity.

Compensation for the released capacity: If there is congestion associated with the released capacity $K - q_1$, the ISO accrues some merchandising surplus from its operation of the stage-2 market. The ISO could return the merchandising surplus to the owners of the physical rights that it has taken possession of, using the difference in nodal prices in the stage-2 market to value these rights. This effectively turns any released physical rights into financial rights. We call this the *use-it-or-get-paid-for-it* rule. Alternatively, the ISO could give the merchandising surplus produced in the stage-2 market to charity or use it to help to defray the ISO's fixed costs. In this case, the holders of the released rights get nothing for them. We call this the *use-it-or-lose-it* rule.

a) *Use-it-or-lose-it rule.* This rule appears to provide the most powerful incentives for physical rights holders not to withhold rights from the market. The release of any rights they withhold to the ISO undermines the profitability of a withholding strategy and they lose entirely the value of any rights withheld from the market that they might otherwise earn if they sold (or used) the rights before the close of the day-ahead market. So, even if there is no free riding and so G_2 holds all the physical rights initially (at the start of stage 1), G_2 does not withhold any. The bid-based dispatch market is inactive. As in section 4.1, it makes a difference whether G_2 can centralize sales to consumers by purchasing power in the North, or whether G_2 sells electricity to consumers without internalizing the value of the rights sold to generators in the North. We thus conclude that, under the use-or-lose-it rule, G_2 obtains the commitment or noncommitment profit corresponding to $q_1 = K$.

Remark: The absence of stage-2 (last day) uncertainty in our model may conceal a potential cost of the use-it-or-lose-it rule if interpreted too rigidly. It may be the case that an a priori efficient plant in the North is scheduled at the end of stage 1 to supply in stage 2 but becomes incapacitated or more generally becomes a high-cost unit at stage 2. Some flexibility should then be created so as to allow substitution possibilities for power at stage 2; the challenge is then how to provide stakeholders with incentives to reallocate production efficiently without altering the spirit of the use-or-lose-it rule. We leave this issue (which does not arise in our model) for future research.

b) *Use-it-or-get-paid-for-it rule*: This rule undermines the direct value to G_2 of withholding physical rights from the market since the withheld rights must be released, but imposes no penalties for withholding rights. Indeed, assuming again the absence of free riding and so G_2 holds all the rights, G_2 in equilibrium withholds all rights and so the bilateral market is inactive. Given that all transmission capacity will be used under any strategy, G_2's total profit (from generation, from the sale of physical rights, and from the dividends received for the financial rights resulting from withheld physical rights) is bounded above by

$$\max_{p_2} \{p_2 D(p_2) - C_2(D(p_2) - K) - K C_1'(K)\}.$$

But G_2 can get exactly this upper bound by withholding all rights and transforming them into financial rights (see section 3). We conclude that the use-it-or-get-paid-for-it rule, while preventing production inefficiency, allows G_2 to optimize against the full demand curve and leads to a high price in the South.

5.2 Surveillance of "gambling" behavior

Our analysis of financial rights demonstrated that in the absence of complete free riding by initial rights holders, G_2 will acquire some financial rights. Note, however, that when the capacity constraint is binding and the financial rights are valuable is also when p_2 will be greater than p_1. This implies that G_2 effectively takes a gambling rather than a hedging position, and welfare is reduced because its ownership of transmission rights increases prices.[21] This necessarily raises the question of whether regulatory oversight can mitigate this source of inefficiency. *One might consider, for example, preventing firms with local market power from buying transmission rights when they involve gambling rather than hedging. That is, positions whose value covaries positively with the value of the player's position in the absence of these rights would be prohibited.*

While a regulatory rule built around this basic principle is likely to provide a useful conceptual framework for designing regulatory surveillance programs, there are several practical problems in applying it in practice in a way that increases welfare. First, in an environment in which there is uncertainty about the nodal price differential, the implications of the previous discussion is that the generator in the South may take on more risk ("underhedge") than it would if purchasing financial rights did not enhance its market power. While one can prohibit taking gambling positions, it becomes difficult to assess the extent of underhedging in practice, though one might be able to do so in specific circumstances.[22]

Second, in our model, one possible benchmark for measurement of "gambling behavior" by

[21]The same property holds for a consumer with market power at the cheap node (see section 7).

[22]The regulatory strategy would then be similar to that adopted for the England-Wales system in the early l990s when the two largest generators created from the existing CEBG were forced to sell a substantial fraction of their output forward using contracts for differences (contracts for differences are insurance contracts against variations in the spot market price). For theoretical analyses of the impact of contracts for differences on market power in systems with uniform prices, see Green (1992), Allaz-Villa (1993) and Allaz (1992). Such analyses in turn are closely related to the Coase conjecture (see. e.g., Tirole 1988).

a player with local market power is that this player's generation profit (or consumption surplus, see Section 7) is positively correlated with the value of congestion rents. So, a simple rule might be that generators are not allowed to acquire financial rights if their value is positively correlated with the value of the firm's generation profits. However, it can be readily demonstrated that this simple rule may not be appropriate. For example, consider a case where supplies available in the North and the South are both uncertain, are highly correlated, and the supplies in the North are large compared to the supplies in the South.[23] For example, there may be hydroelectric supplies (none of which belonging to G_2) at both locations, and the amount of energy they can produce is contingent on the same random variations in rainfall from year to year. G_2's generation profit is then negatively correlated with rainfall. If furthermore there is much more hydroelectric supply in the North than in the South, p_2 varies much more than p_1 and the value of financial rights is positively correlated with rainfall. Thus, from a statistical viewpoint, G_2 can hedge by purchasing financial rights. But at the same time we showed above that such purchases increase its market power. This obviously complicates the regulator's surveillance problem.

Another surveillance index might be to examine whether it can be demonstrated that the acquisition of financial rights by a local monopoly generator led to an increase in the *difference* in nodal prices. In our analysis above, the voluntary purchase of financial rights by players with local market power both increased the value of these rights (by increasing the difference in nodal prices) and reduced output and economic welfare. The two effects do not always go together, however. It can be shown that with other market power configurations the purchase of financial rights by players with market power may increase welfare and congestion rents simultaneously.[24]

Accordingly, regulatory surveillance of transmission rights ownership that turns on "gambling" or "underhedging" behavior is likely to be difficult to implement under many realistic

[23] A formal example of this phenomenon is worked out in Joskow and Tirole (1999a).

[24] A formal example is developed in Joskow and Tirole (1999a). Our example builds on Stoft's (1997, 1998) analysis of strategic expropriation of the rights' value by a generator in the North. By adding an inefficient fringe in the North to his analysis, we can show that the expropriation does not just lead to a redistribution of wealth, but also to a reduction in welfare if the inefficient fringe supplies in equilibrium.

real world supply situations.

6 Financial vs physical rights: welfare comparison

6.1 Welfare comparison

We now compare physical and financial rights for the market power configuration that we have focused on from a welfare perspective, examining the effects of rights allocations on market power and production inefficiency. Financial rights can enhance G_2's market power, but do not lead to production inefficiency. Physical rights can both enhance G_2's market power and lead to production inefficiency. However, both the market power enhancing effects of physical rights and the production inefficiency effects depend on the ability of G_2 to commit to an ex post supply strategy. We examine the welfare associated with four sets of cases:

Case 1: Physical rights without capacity release under commitment.

Case 2: Financial rights; or physical rights with either a use-or-get-paid-for-it rule; or a use-or-lose-it rule with commitment.

Case 3: Physical rights without capacity release with noncommitment.

Case 4: No rights; or physical rights with a use-or-lose-it rule and noncommitment.

Letting W_i and Π_i denote social welfare and G_2's profit in case i, the appendix shows that:

$$\Pi_1 \geq \max(\Pi_2, \Pi_3) \geq \min(\Pi_2, \Pi_3) \geq \Pi_4,$$

$$W_4 > W_2 \geq W_1 \text{ and } W_4 > W_3.$$

The welfare comparisons are displayed in Figure 2. Figure 2 assumes away free riding and therefore posits that gains from trade between the generator with market power and the rights owners are realized. In Figure 2, welfare decreases when moving east (increase in local market power) or north (increased withholdings).

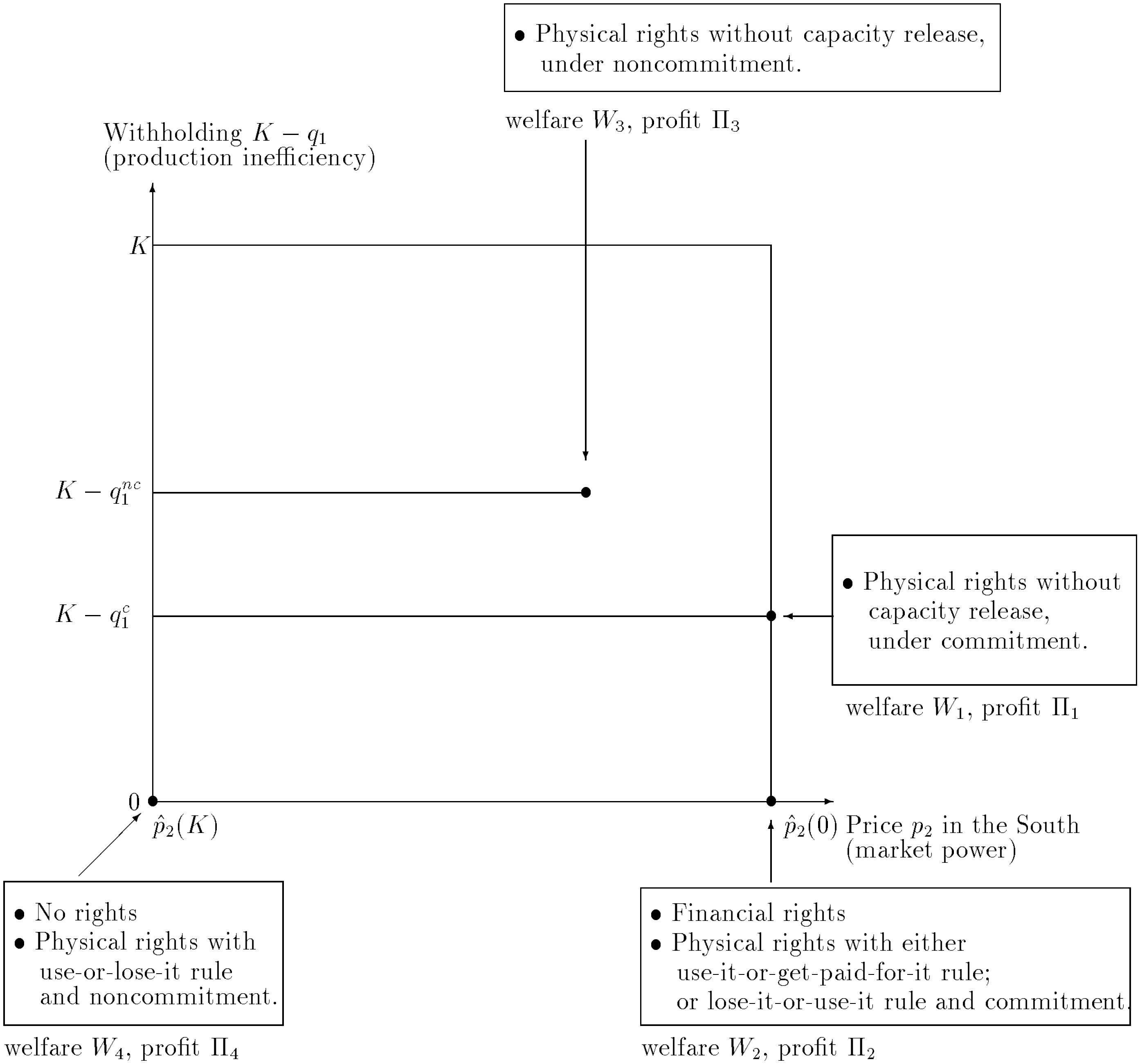


Figure 2

6.2 Discussion of welfare comparisons

A striking implication of our policy analysis is that the absence of transmission rights (the "zero net supply solution") does as well as and in general better than either system of rights. This leads naturally to the question of why is the ISO creating financial or physical rights if insurance opportunities can be created "synthetically" through ordinary insurance markets or contracts for differences. There may be two reasons why a positive net supply may be unavoidable; we have not explored either reason and think this topic is a central area of potential research in view of the fact that all current policy proposals emphasize institutions with positive net supply of rights.

First, it may be the case that zero net supply (pure insurance markets) is not a feasible option. The ISO's merchandizing surplus must go to someone. To the extent that it goes to nonstakeholders or to stakeholders with no local market power, how can we prevent side deals between these investors and large stakeholders, that is stakeholders like G_2 who through their local market power can affect the value of the rights? Avoiding such side deals requires some form of "insider trading regulation", in which stakeholders with market power are not allowed to engage in side deals. The question then is: If one can prevent such side deals under zero net supply, can't one also prevent perverse holdings of financial rights by large stakeholders under positive net supply (see the discussion of the prohibition of "gambling behaviors" in section 5.2)? We leave this issue for future research.

Another argument may be that the creation by the ISO of transmission rights is required for the provision of transmission investment incentives. According to Hogan (1992), when new transmission investments are made, the ISO is supposed to create new financial rights to match the additional network capacity that has been created by the new transmission investments. The dividends from these financial rights then are supposed to become the (sole) source of the transmission investors' revenue. A similar investment motivation is associated with physical rights. The study of long-term incentives for investments in transmission is still in its infancy, and much work will be required in order to understand the articulation between these incentives and the design of transmission rights.

7 Alternative market power configurations

We now discuss whether and how physical and financial transmission rights enhance market power with several alternative market power configurations. This discussion is designed to be intuitive and illustrative. See Joskow-Tirole (1999a,b) for more detail.

7.1 Generator market power at the cheap node

Suppose that G_1 is a monopoly in the North, while production in the South is competitive and there is no buyer market power.

a) *Financial rights*: In contrast with the case of generator market power at the expensive node, financial rights holding by the generator with market power at the cheap node has no impact. To see this, suppose first that G_1 holds no financial rights. As is well-known from the literature (e.g., Oren 1997, Stoft 1999), financial rights are then worthless: Financial rights on the line of a two-node network can have positive value only if the capacity is fully used. Suppose that $q_1 = K$ and $p_1 < p_2$. Then G_1 can bid to supply a fixed amount $K - \varepsilon$ (where ε is small) and raise p_1 discontinuously to p_2 (which is hardly affected).

In a sense, G_1 effectively already "owns" the transmission rights even if formally owns no financial rights. So G_1's holding financial rights has no effect on prices or quantities and does not enhance its market power.

Remark 1: If there is a monopoly in the North as well as in the South, then the monopoly in the North will capture all of the congestion rents (by bidding a fixed quantity $q_1 = K - \varepsilon$ for ε substantially small) and the value of financial rights will be zero. Thus, adding a monopoly generator in the North when there is a monopoly generator in the South *mitigates* the market power enhancing effect of financial rights on the market power of the monopoly generator in the South and leads to *lower* prices p_2 in the South compared to the case when there is competitive generation in the North. That is, even if the monopoly generator in the South holds a fraction α_2 of financial rights, the equilibrium price in the South is $p_2(0) = p_2^m$ rather than $p_2(\alpha_2)$.

Remark 2: The result that ownership of financial rights has no effect on the behavior of the monopoly generator in the North does not carry over to all cases where there is imperfect competition (market power) in the North. For example, when there is a dominant generator in the North plus a competitive fringe in the North, allocating financial rights to the dominant generator can lead to lower prices and eliminate production inefficiency associated with its "no rights" strategic behavior. The financial rights give the dominant generator a less costly way of extracting congestion rents.[25,26]

b) *Physical rights*: Suppose that the monopolist in the North acquires all rights from the non-stakeholder owner (no free-riding). G_1 may withhold some rights and use the others to dispatch its supplies. Note, though, that in the case of constant marginal cost in the South there is no withholding by G_1 at all ($q_1 = K$); for, if G_1 signs q_1 contracts with consumers in the South, then competitive behavior of generators in the South yields p_2 such that

$$p_2 = C_2'(D(p_2) - q_1),$$

and G_1 chooses q_1 so as to maximize

$$q_1 p_2 - C_1(q_1),$$

where p_2 is given by (1). So, *if* $q_1 < K$ then

$$C_2' - C_1' = q_1[C_2''/(1 - C_2''D')].$$

This is impossible if $C_2'' = 0$. In contrast, if marginal cost is upward sloping in the South it may be profitable for G_1 to withhold output ($q_1 < K$) in order to raise the price in the South. Note, though, that physical rights are not needed by G_1 to implement this withholding strategy. Under financial rights, G_1 captures congestion rents ($p_1 = p_2$) and schedules the same profit maximizing value of q_1 with the ISO as under physical rights.

[25] This case is worked out in detail in Joskow-Tirole (1999a), pp24-25.

[26] Financial rights holdings by the monopoly generator in the North in general also enhances welfare if power is consumed in the North: Introduce a demand $D_1(p_1)$, and for computational simplicity assume that the price p_2 in the South is determined by a constant returns to scale technology located there ($p_2 = c_2$). If G_1 does not hold rights, then G_1 selects p_1 so as to solve

$$\max_{p_1 \le c_2} \{p_1[D_1(p_1) + K] - C_1(D_1(p_1) + K)\}.$$

In contrast, G_1 solves

$$\max_{p_1 \le c_2} \{p_1[D_1(p_1) + K] - C_1(D_1(p_1) + K) + (c_2 - p_1)K\},$$

when holding all rights. The price in the North is always weakly smaller when G_1 holds the rights.

7.2 Consumers with monopsony power at the expensive node

a) *Financial rights* : Consider the case where there is buyer market power (a monopsony) in the South. If the monopsony holds financial rights, its monopsony power actually decreases since the value of financial rights declines as the price p_2 in the South declines. [This does not mean that the monopsony will not acquire the rights; after all, its demand behavior impacts the value of rights, and so its acquiring rights help internalize this externality and creates gains from trade.]

b) *Physical rights* : The monopsonist in the South purchases all rights from the nonstakeholder owner, and then purchases q_1 units of power in the North at price $p_1 = C_1'(q_1)$, and q_2 units of power in the South at price $p_2 = C_2'(q_2)$. Denoting by $S(q_1 + q_2)$ the monopsonist's gross surplus, the latter maximizes $\{S(q_1+q_2) - q_1 C_1'(q_1) - q_2 C_2'(q_2)\}$ over input purchases $\{q_1, q_2\}$. If returns in the North are constant or do not decrease fast (C_1'' small), then there are no withholdings ($q_1 = K$) and the outcome is the same as under financial rights.[27] If C_1'' is large, the monopsonist withholds rights ($q_1 < K$) to extract rents from the generators in the North; higher rights prices then more than compensate for the effects of reduced supplies from the North on the price in the South. The monopsonist and the generators in the South are better off with physical rights than under financial rights, and the generators in the North are worse off.[28]

7.3 Consumers with monopsony power at the cheap node

a) *Financial rights* : If there were a monopsony in the North and competitive behavior in the South, the behavior of the monopsony in the North could be affected if it held financial rights. By reducing the price in the North (which is feasible if marginal cost in the North

[27] Under financial rights, the generators in the North are dispatched, and there is no way for the monopsonist in the South to capture their inframarginal rents. The monopsonist solves:

$$\max_{q_1} \{S(K+q_2) - (q_2+K)C_2'(q_2) + [C_2'(q_2) - C_1'(K)]K\}$$
$$= \max_{q_2} \{S(K+q_2) - q_2 C_2'(q_2) - K C_1'(K)\}.$$

[28] When $C_1(q_1) = c_1 q_1 + b\frac{q_1^2}{2}$, $C_2(q_2) = c_2 q_2$, and $c_2 - c_1 < bK$ (so there are withholdings), then $q_1 = (c_2 - c_1)/b$ and $S'(q_1+q_2) = c_1$ under physical rights. So total output is the same as under financial rights; production inefficiency under physical rights implies that financial rights dominate physical rights from a social welfare perspective.

is strictly increasing), the monopsonist increases the difference between the nodal prices and the congestion rents it receives. This leads to further distortion of demand in the competitive upstream market (compared to monopsony without rights) and a reduction in welfare. So, if there are consumers with market power in the exporting region, allocating to them financial rights increases their incentives to reduce purchases, enhances their market power and reduces welfare.[29]

b) *Physical rights* : Again, the potentially interesting twist associated with physical rights involves the potential for the allocation of physical rights to restrict exports from the North. Under what if any conditions would a monopsony buyer in the North benefit (net of the cost of the rights) by acquiring and then withholding physical rights from the suppliers in the North to further reduce the nodal price in the North by restricting exports?[30] If the marginal cost curve in the North is upward sloping, the large consumer in the North may indeed have an interest in withholding rights. As in the case of monopoly power in the South, it is important to account for the Coasian problem, if any. Suppose that there is no free riding, and so the monopsonist in the North acquires all physical rights. And assume first that this monopsonist is allowed to purchase electricity from G_1 and resell it to consumers in the South. Because the monopsonist holds all the rights, the monopsonist then also enjoys an export monopoly in this *commitment case* . It can be shown that the monopsonist withholds some rights (i.e., does not use them to export power) under basically the same condition as the monopoly in the South. Assuming for simplicity that marginal cost in the South is constant, here *the condition for withholding is that the difference in marginal costs between South and North be smaller than total production in the North times the increase in the marginal cost in the North associated with a unit increase*

[29]The parallel between a monopoly at the import node and a monopsony at the export node is not fortuitous. As long as the link is congested, the monopsony in the North is mathematically equivalent to a monopoly in the South, provided that a) demands are treated as negative supplies, and b) the residual demand curve in the South $(D(p_2) - K)$ is replaced by the residual supply curve $(S_1(p_1) - K)$ in the North.

[30]Obviously, buyers in the North would be very interested in convincing the government to restrict exports in order to reduce local nodal prices. So, we should not be surprised to find consumer groups in exporting areas to be cautious about deregulation and increased exports from their low-cost suppliers. Since there are gains from trade, it would make more sense for regulators to give local consumers an entitlement to a share of the additional profits earned from price deregulation and unrestricted exports (e.g. regulatory entitlements to export profits), rather than restricting exports of cheap power.

in production.[31]

When the monopsonist in the North is not allowed to export power (the *noncommitment case*), the monopsonist faces the same problem as Coase's durable good monopolist: After having sold the rights to generators in the North, it no longer internalizes the increase in the value of the rights associated with a contraction of demand in the North. The monopsonist thus consumes more energy than in the commitment case, which per se increases welfare. However this Coasian problem also provides the monopsonist with increased incentives to withhold rights as a way to "commit" not to consume much in the power market.

Accordingly, buyers located in an exporting region may try to exploit a physical rights system by engaging in collective action to withhold export rights in order to drive the local price for power down below competitive levels.

7.4 Oligopolistic competition

We have assumed local monopoly (or monopsony) power. Our analysis can be generalized to oligopolistic competition at a node. For example, there could be several generators in the importing region. In this case and in the absence of rights, the standard oligopoly analysis for unconstrained networks (Klemperer-Meyer (1989), Green (1992), and Green-Newbery (1992)) applies verbatim to the residual demand curve in the South. Hence, each generator in the South has some market power.[32] As in the local monopoly case, a generator in the South has the ability to raise the price in the South and its incentive to do so is enhanced by his holding financial rights. Similarly, this generator may benefit from the withholding of physical rights.

While the oligopoly case differs in degree rather than in nature from the monopoly situation, the oligopoly case introduces an interesting new twist: Jacking up the price in the South (or lowering it in the North) is a "public good" for the community of producers in the South: All

[31] Let q_1 denote the level of exports and b_1 the monopsonist's consumption. So total production in the North is $q_1 + b_1$. Letting $\mathcal{S}(b_1)$ denote the monopsonist's gross surplus, then the monopsonist maximizes

$$\max_{\{b_1, q_1 \leq K\}} \{\mathcal{S}(b_1) - (q_1 + b_1)C_1'(q_1 + b_1) + c_2 q_1\}.$$

The derivative with respect to q_1 is equal to

$$c_2 - C_1' - (q_1 + b_1)C_1''.$$

[32] Klemperer and Meyer show that the outcome lies between Bertrand and Cournot competition if there is enough uncertainty. Of course, the extent of market power would increase if the generators engaged in repeated game strategies.

want this public good to be supplied, but each would like to free ride on output curtailments and thus on financial rights ownership (or withholdings of physical rights) by the other oligopolists. The study of this public good game follows the standard lines.[33]

Another interesting point is that it matters whether financial rights holdings are made public before the energy market opens. Under secret holdings, oligopolists always buy some financial rights (in the absence of full free riding by investors). In contrast,[34] public holdings introduce a strategic effect and may result in the oligopolists not holding any rights: A generator in the South which purchases rights shifts its reaction curve toward lower quantities and encourages its rival(s) to produce more. This "strategic substitutes" effect may dominate the value-of-rights-enhancement effect; furthermore, it is absent when only one player has market power at a node.

7.5 Summing up

While several cases must be considered depending on who has market power, the general logic is simple and intuitive: Financial rights holdings by a producer in the importing region or by a consumer in the exporting region aggravates their market power since financial rights give them an extra incentive to curtail their output or demand to make the rights more valuable. In contrast, financial rights holdings by a monopsony in the importing region mitigate its market power by giving it an incentive to raise price in the importing region. Finally, financial rights holdings by a monopolist in the exporting region may have no impact on market power, since the monopolist may already be able to capture the congestion rents in the absence of rights.

The results for physical rights are similar to those for financial rights. There are two primary differences. First, a physical rights system introduces the possibility that owners of transmission rights can withhold these rights from the market, effectively reducing the capacity of the constrained transmission link. Second, the ability of generators to commit to ex post

[33]We are very grateful to a referee for fully solving out this public good game in the case of two generators in the South constrained to offer vertical supply schedules (that is, they announce inelastic quantities and thus play a Cournot game).The referee further showed that in a situation in which one of the duopolists in the South moves first in the energy market, this Stackelberg leader purchases more financial rights than the Stackelberg follower, since he benefits more than his rival from a price increase.

[34]As pointed to us by one of the referees.

supply strategies affects the impact of rights on market power and the extent of withholding and the associated supply-side inefficiency. In the case of ex post supply commitment, generators or consumers with market power in the importing region can use physical rights to their advantage by driving down the prices received by generators in the exporting region. So too can consumers with market power in the exporting region. They are all better off than under a financial rights system, but withholding leads to supply-side inefficiency. A generator with market power in the North does not need physical rights to withhold output. If ex post supply commitment is not possible, a monopsonist holding such rights in the North would withhold more, increasing supply inefficiency as is the case for a generator with market power in the South.

8 Loop Flow Considerations

While some networks or subnetworks are well approximated by the radial structure that we have studied until now, more complex networks exhibit loop flows associated with the fact that electrons follow the path of least resistance. On electrical networks with multiple interconnected links, the patterns of electricity flows follow physical laws known as Kirchhoff's laws (Schweppe et al. 1988). For example, in a three-node network (e.g., Figure 3), a power injection at one node (e.g., node 1 in Figure 3) and an equal amount withdrawal at another node (e.g., node 3 in Figure 3) affect not only the congestion on the line linking the two nodes, but also the congestion on the other two lines (node 1 to node 2 and node 2 to node 3) as well. The distribution of the power flows over these multiple interconnected links is simply referred to as "loop flow." This section shows that many of the results regarding the impact of transmission rights on generator market power for a two-node network carry over in the presence of loop flows.[35] The primary differences arise as a consequence of the complementarities between generators at different nodes created by loop flow. Following an extensive literature on the topic,[36] we illustrate the ideas by means of a simple three-node network, described in Figure 3, with two nodes of (net) production and one node of (net) consumption. We ignore losses

[35]We consider here only the anology to the two-node case with generator market power in the South (expensive node) and competitive generation in the North (cheap node).

[36]See, e.g., Nasser (1997), Oren et al (1995), and Oren (1996).

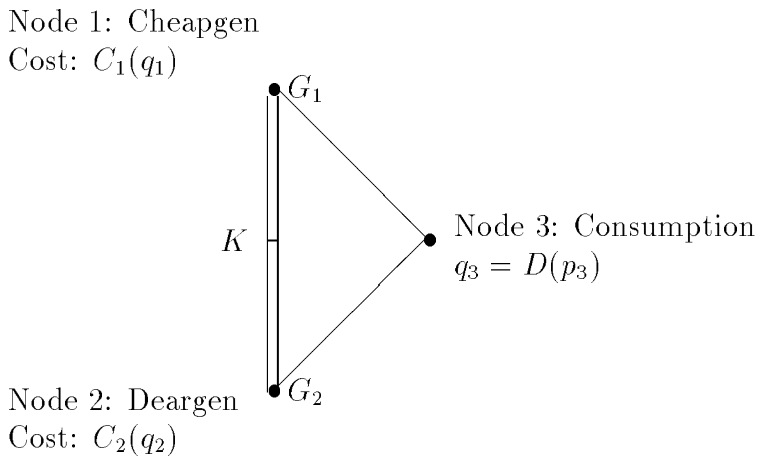


Figure 3

and assume that the three links have equal length and impedance. We focus on the case in which the transmission line between the two generation nodes is congested, which is the direct analog of the two-node network and then discuss briefly how the analysis changes for different patterns of congestion.

As in the two-node network, nodes 1 and 2 are the low- and high-marginal- cost production nodes respectively ($C_1' < C_2'$ over the relevant range). Furthermore, production is competitive at node 1 and produced by a monopoly G_2 at node 2. Consumption occurs now at node 3. There, consumers demand $q_3 = D(p_3)$, with inverse demand function $p_3 = P(q_3)$. The line between nodes 1 and 2 has capacity K and is congested, while the other two lines have excess capacity.

Since we assume that there are no losses on the lines,

$$q_1 + q_2 = q_3. \tag{3}$$

The physical laws of electricity (Kirchhoff's) determine the flows through the three transmission lines. A reader unfamiliar with electric networks can think about electricity balance in the following terms: Electricity flowing from node $i \in \{1, 2\}$ to node 3 must follow the path of least resistance. This implies that the "resistances" encountered along the two possible paths (direct, and indirect through the other production node) are equal.[37] Because the indirect path is twice as long as the direct path, so is the "resistance". This implies that a unit production at one node generates a one-third flow through along the indirect path and twice as much along the direct path.[38] Because we focus here on the congestion of the line between the two generating nodes, the constraint becomes

$$\left| \frac{q_1}{3} - \frac{q_2}{3} \right| \leq K,$$

or using the fact that there is a lower marginal cost and so more production in the North than in the South[39]

$$q_1 - q_2 \leq 3K. \tag{4}$$

8.1 Bid-based dispatch

8.1.1 Absence of financial rights

Under bid-based dispatch, the ISO dispatches productions, q_1 and q_2, and consumption q_3 so as to maximize social surplus (consumer surplus minus total production cost, as revealed by the players) subject to the feasibility constraints $\{(3), (4)\}$.[40] Letting η denote the shadow price of constraint (4), the nodal prices satisfy at the optimum:

$$p_1 = p_3 - \frac{\eta}{3}, \tag{5}$$

[37] The difference in phase (the analogy of voltage for AC networks) is equal to the product of impedance (resistance) and power flow (current).

[38] These flows may be fictitious. Indeed, if generators at nodes 1 and 2 produce the same output and therefore generate the same, but opposite fictitious flows through the line located between them, no power actually flows through that line since the two fictitious flows cancel.

[39] It is natural to ask why there is a transmission link at all between nodes 1 and 2. Why would a transmission link be built between nodes 1 and 2 rather than adding capacity to the link between node 1 and node 3 so that the cheap power could be delivered directly to consumers at node 3? Alternatively, given that the link has been built for some reason it could be that it would be beneficial (at least to G_1) to close down the link between nodes 1 and 2 so that all of the cheap power can flow directly to consumers at node 3. Transmission links like the one between node 1 and node 2 are typically built for reliability reasons. For example, one of the other links may fail and would be unable to support supplies directly from one of the generating nodes to node 3. The link between nodes 1 and 2 provides an alternative delivery path. More generally, the configuration of the transmission network in the long run is endogenous and reflects investment decisions involving both generation and transmission. As we indicated at the outset, we are examining electricity networks with fixed capital stocks and leave these interesting investment issues to future research.

[40] See Schweppe et al. (1988) for the generalization of the following analysis to an arbitrary electrical network.

and

$$p_2 = p_3 + \frac{\eta}{3}. \tag{6}$$

And so

$$p_3 = \frac{p_1 + p_2}{2}. \tag{7}$$

To understand (5) and (6), note that in the absence of a transmission constraint everything would be produced at node 1, since production there is both competitively supplied and cheaper than production at node 2. A congested line limits the production at node 1. An extra unit produced at node 1 generates an added load of one-third on the congested line and so must be subject to a "tax" equal to $\eta/3$ (i.e., one third of the shadow price of the constraint). Conversely, a unit production at node 2 unloads the constraint by one third and should therefore receive a "subsidy" equal to $\eta/3$. The "tax" paid by generators at node 1 is the price consumers pay at node 3 less 1/3 of the shadow price of congestion ($\eta/3$). That is, the equilibrium price generators at node 1 see is less than the equilibrium price consumers pay at node 3. The "subsidy" provided to generators at node 2 is the price consumers pay at node 3 plus 1/3 of the shadow price of congestion. That is, the equilibrium price generators see at node 2 is greater than the equilibrium price consumers pay at node 3.

Using (3), (4) and (7), we thus obtain:

$$p_2 = 2P(2q_2 + 3K) - p_1.$$

Competitive behavior at node 1 further implies that

$$p_1 = C_1'(q_1) = C_1'(q_2 + 3K).$$

We therefore obtain the profit from generation for the monopoly at node 2 (which we here write for convenience as a function of q_2 rather than p_2):

$$\mathcal{G}(q_2) \equiv p_2 q_2 - C_2(q_2),$$

or

$$\mathcal{G}(q_2) \equiv \left[2P(2q_2 + 3K) - C_1'(q_2 + 3K)\right] q_2 - C_2(q_2). \tag{8}$$

In the absence of financial rights, G_2 selects $q_2 \geq 0$ so as to maximize $\mathcal{G}(q_2)$. It is interesting to note that G_2 receives two benefits from withholding output in comparison with the case where it is price taker. Recall that $p_2 = p_3 + (\eta/3)$. By reducing q_2, G_2 increases the consumer price $p_3 = P(2q_2 + 3K)$; actually the standard contractionary effect (the elasticity of demand) is doubled since, unlike the two-node case, a reduction in the production in the South forces an equal reduction of the output in the North. In other words production in the South and the North are *local complements*, where "local" refers to the fact that for large transmission capacities the two outputs become substitutes. Second, a reduction in output q_2 increases congestion and thereby the subsidy $\eta/3 = p_3 - p_1$ received for production in the South. [This subsidy increases because p_3 increases and also, if production in the North exhibits decreasing returns to scale, because the cost of the marginal plant in the North decreases, making q_2-enabled production in the North more desirable.]

8.1.2 Financial rights

With more than two nodes, there are at least two ways of introducing financial rights:

• *Link-based rights*: Link-based rights are financial rights associated with a transmission line and paying a *dividend equal to the shadow price of the congestion on that line* . Such rights are for example being put in place in California associated with inter-zonal congestion.

In our context, the only link-based rights with positive value are those attached to the link between nodes 1 and 2. Suppose that K such rights are issued; then the total dividend is ηK. This total dividend corresponds exactly to (and therefore can be covered by) the merchandizing surplus, that is

$$p_3q_3 - p_1q_1 - p_2q_2 = (p_3 - p_1)q_1 + (p_2 - p_1)q_2 = \frac{\eta}{3}(q_1 - q_2) = \eta K$$

(since $q_1 - q_2 = 3K$).

• *Fictitious-bilateral-trades-based rights*: Financial rights were first designed by Hogan (1992) in a different way. He considered the set of bilateral trades that are feasible (meaning here that they satisfy (7) and (8)). Such trades may, but need not be those that actually occur.

For example, suppose that producers at node i $(i = 1, 2)$ fictitiously sell $\bar{q}_i$ units in bilateral contracts with consumers at node 3. Those trades are feasible if $\bar{q}_1 - \bar{q}_2 \leq 3K$. These fictitious trades create by definition $\bar{q}_1$ financial rights between *nodes* 1 and 3, yielding dividend $p_3 - p_1$ each, and $\bar{q}_2$ financial rights between *nodes* 2 and 3, yielding dividend $p_3 - p_2$ each, where prices refer to the ex post equilibrium prices for the market outcomes (and thus not necessarily to the prices corresponding to the fictitious trades). The total dividend to be paid for prices (p_1, p_2, p_3) is therefore

$$(p_3 - p_1)\bar{q}_1 + (p_3 - p_2)\bar{q}_2 = \frac{\eta}{3}(\bar{q}_1 - \bar{q}_2) \leq \eta K.$$

Thus the dividend can again be covered by the merchandizing surplus.[41]

Without loss of generality, we here consider link-based rights. Suppose that K such rights are held by a nonstakeholding investor who then (as in section 3.1) resells them to the monopoly producer in the South. The value of these rights is

$$\mathcal{F}(q_2) = \eta K = 3[P(2q_2 + 3K) - C_1'(q_2 + 3K)]K.$$

The value of financial rights decreases with q_2 for the now familiar two reasons: decrease in consumer price and increase in the marginal cost in the North due to local complementarity.

G_2 thus solves

$$\max_{q_2} \left\{ \mathcal{G}(q_2) + \mathcal{F}(q_2) \right\},$$

and, as in the radial network case, restricts output further than in the absence of financial rights. Indeed, the result according to which G_2 optimizes against the *full* demand curve generalizes, since

$$\begin{aligned} \mathcal{G}(q_2) + \mathcal{F}(q_2) &= p_3 q_3 - p_1 q_1 - C_2(q_2) \\ &= P(2q_2 + 3K)(2q_2 + 3K) - q_1 C_1'(q_1) - C_2(q_2). \end{aligned}$$

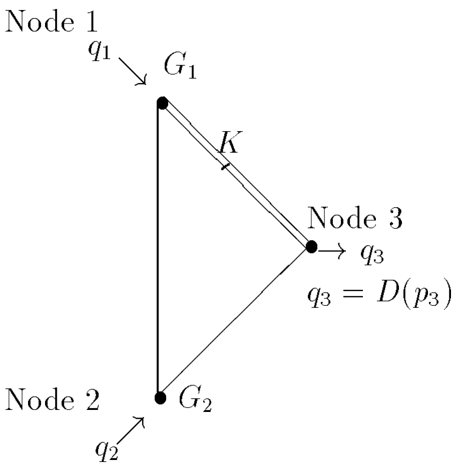
With the required adjustments, the insights of section 3 thus carry over from the two-node network.[42]

[41] This result extends to arbitrary networks: see Hogan (1992) and the appendix of Chao-Peck (1996). The two types of rights are equivalent as long as a) the amount of link-based rights is equal to the capacity of the line, and b) the fictitious trades in the Hogan approach exhaust the transmission constraint.

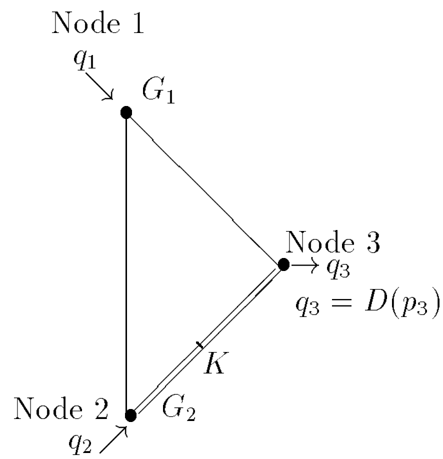
[42] Bill Hogan pointed out to us, though, that the point that a monopoly generator at the cheap node fully

8.1.3 Other patterns of congestion

Let us now assume that one of the direct links between a production node and the consumption node is congested while the other links are not: see Figure 4. We keep the notation "K" for the capacity of the congestion link even though the identity of that link has changed.



(a) G_2 on the good side of the constraint



(b) G_2 on the wrong side of the constraint

Figure 4

In case (a), the capacity constraint on the line between nodes 1 and 3 can be written as

$$\frac{2}{3}q_1 + \frac{1}{3}q_2 \leq K,$$

where K now denotes the capacity of that line. Note that the two outputs are now *local substitutes*. A contraction in output (q_2) at the expensive node leads to an increase in output

appropriates the congestion rent (and so financial rights holdings have then no impact) does not carry over to the loop-flow case. In the two-node situation with a competitive supply in the South, say, a monopoly G_1can bid a quantity just below the capacity K of the link and make it look uncongested. In our three-node network, this strategy leads to a slightly smaller quantity produced at node 2, and so the link still appears congested.

The difference between the two- and three-node networks is, we feel, more quantitative than qualitative; for, the presence of uncertainty in the two-node network would invalidate the irrelevance of financial rights holding by a monopoly G_1. If the exact capacity of the link were unknown at the time G_1 bids, it would in general be suboptimal for G_1 to fully expropriate the congestion rent with probability one. And then G_1's holding of financial rights would reduce his incentive to expropriate the congestion rent, just like in the three-node network. See also Remark 2 in section 7.1.

(q_1) and an associated increase in the marginal cost at node 1. The reader can check that[43]

$$\mathcal{F}(q_2) = \left[\frac{P\left(\frac{3K+q_2}{2}\right) + C_1'\left(\frac{3K-q_2}{2}\right)}{2}\right] q_2 - C_2(q_2),$$

and

$$\mathcal{G}(q_2) = \frac{3}{2}\left[P\left(\frac{3K+q_2}{2}\right) - C_1'\left(\frac{3K-q_2}{2}\right)\right] K.$$

Again, G_2 optimizes against the full demand curve (that is, maximizes $p_3q_3 - q_1C_1'(q_1) - C_2(q_2)$), when owning the financial rights. The impact of G_2's ownership of financial rights is a priori more ambiguous than in the two-node network. An output contraction by G_2 raises the consumer price, but also by a substitution effect, increases the marginal cost in the North. The net effect on the shadow price of the congested line is a priori unclear, unless returns to scale in the North decrease slowly, in which case G_2's ownership of financial rights enhances market power. If the marginal cost of the generators in the North were constant, allocating financial rights to G_2 would lead it to contract output further than it would in the absence of financial rights.

Case (b) is more interesting. There, G_2 is on the wrong side of the capacity constraint, and does not produce in the absence of financial rights: Its marginal cost is higher than that of producers in the North, and furthermore G_2 makes twice as much use of the congested line as they do and therefore gets taxed twice as much.[44]

Suppose now that G_2 owns the financial rights. Despite its double disadvantage, G_2 may produce so as to enhance the value of financial rights. Straightforward computations show that

$$\begin{aligned}\mathcal{F}(q_2) + \mathcal{G}(q_2) &= p_3q_3 - q_1C_1'(q_1) - C_2(q_2),\\ &= P(3K-q_2)(3K-q_2) - (3K-2q_2)C_1'(3K-2q_2) - C_2(q_2).\end{aligned}$$

[43]Letting η denote the shadow price of the congested line, the nodal prices are

$$p_1 = p_3 - \frac{2\eta}{3} \text{ and } p_2 = p_3 - \frac{\eta}{3}.$$

[44]The capacity constraint is now

$$\frac{q_1}{3} + \frac{2q_2}{3} \leq K,$$

and the nodal prices are

$$p_1 = p_3 - \frac{\eta}{3} \text{ and } p_2 = p_3 - \frac{2\eta}{3}.$$

and so

$$\left.\frac{d\left(\mathcal{F}(q_2)+\mathcal{G}(q_2)\right)}{dq_2}\right|_{q_2=0} \equiv -3P'(3K)K + 6KC_1''(3K) + \left[2C_1'(3K) - P(3K) - C_2'(0)\right].$$

The term in brackets in the right-hand side of the derivative is approximately equal to 0 if i) the line is hardly congested when G_2 does not produce,[45] and ii) $C_2'(0)$ is close to $C_1'(3K)$. Under i) and ii), G_2's two handicaps relative to G_1 are small, and so G_2 gains by increasing its load and making the line appear more congested.

This artificial loading of the line by its owner is reminiscent of the example given in section 7.1, in which the monopoly owner in the North (thus on the wrong side of the constraint in the two-node network) has an incentive to increase its supply when owning the financial rights on the congested line. The welfare implications of this strategic load are however quite different. Increased supply in section 7.1 eliminated the inefficient fringe in the North and improved welfare. Here, increased supply has two perverse effects : By locating some production near the constraint, it reduces total supply to the consumers; and it substitutes expensive power for cheap power, resulting in production inefficiency.

We have not examined other seller and buyer market power configurations and leave this analysis to future research.

8.2 Physical rights: Additional considerations raised by loop flow

Many of the similarities and differences between physical and financial rights identified for a two-node network carry over to a three-node network. Accordingly, we will not repeat them here. We examine here only the standard loop flow problem described in Figure 3 (only the line between nodes 1 and 2 is constrained) in order to understand *additional* institutional issues that arise with physical transmission rights rather than financial rights in the presence of loop flow.

We make three observations regarding physical transmission rights on a network with loop flow. Even in the absence of market power associated with the production or purchasing of

[45] That is, $p_3 = P(3K)$ is close to $p_1 = C_1'(3K)$.

electricity, the efficient implementation of a physical rights system on a network with loop flows must confront a number of significant challenges. These challenges must be understood to talk intelligently about physical rights systems for managing congestion on electric power networks.

Observation #1: Imputing transmission capacity usage to a bilateral contract under loop flows.

Because an injection at one node of the network and an equal withdrawal at another node affect the flows through all links, the ISO must verify that the players scheduling a bilateral trade also possess the relevant physical rights on the network's links. For example, for our simple three-node network, a generator in the North (node 1) selling 1MW to a consumer at node 3 must own two thirds of a physical right on the line from node 1 to node 3, and one third of a physical right on the lines from node 1 to node 2 *and* from node 2 to node 3.

The designer of a physical rights system a priori can choose between two types of rights accounting systems: *a system with an exhaustive set of bidirectional rights or a system with a parsimonious set of unidirectional rights.* In the former case, the designer creates six rights, that is one per line in each direction. In the latter case, the designer contents herself with three directed rights (one per line), *and* allows for negative capacity usage. For example, when selecting directed rights from 1 to 3, 1 to 2, and 2 to 3, then a bilateral unit trade between G_2 and a consumer at node 3 consumes two thirds of a unit of transmission capacity on the 2-3 line (direct path), *minus* one third on line 1-2 and *plus* one third on line 1-3 (indirect path). We will discuss shortly the feasibility of either approach.

Observation #2: Unloading a link: creation of rights vs netting.

Ignoring for the moment market power, that is both G_1 and G_2 produce competitively, a fundamental issue in a physical rights system with loop flows relates to the provision of incentives for a generator located in the South to unload the congested link.

In the *exhaustive* set of bidirectional rights case, 5 out of the 6 types of rights are valueless provided that the corresponding directed flows do not congest their respective lines. Only physical rights for capacity for transferring power from node 1 to node 2 have positive value, η say. Then, G_2 receives no direct financial incentive (or "subsidy") for unloading the line. A bilateral trade by G_2 with consumers yields G_2 price $p_2 = p_3$ per unit. G_2 should receive $p_2 = p_3 + \eta/3$ to have the proper incentives to produce. In contrast, a bilateral trade between a generator in the North and a consumer yields the generator $p_1 = p_3 - \frac{\eta}{3}$, as it should be. The basic problem here is that the value to generators in the North (G_1) of the generator in the South (G_2) producing some additional output is greater than the cost to G_2 of producing that additional output (note that we continue to assume that G_2 behaves competitively). If G_2 produced more then the G_1 generators could produce more as well. Thus, there is an opportunity for G_2 to enter into mutually beneficial production and sales agreements with the generators at G_1 that would result in G_2 producing more and getting paid more for what it produces. For example, G_2 could contract with generators in the North offering to supply q_2 overall (recall $q_2 < q_1$) and bundle its own output q_2 with theirs to sell $2q_2$ to consumers at node 3. G_2 would then get implicit credit for the value of its unloading the congested line by $q_2/3$. Netting[46] would occur as long as the ISO recognizes that there is no net flow created by the bundled outputs along the congested line, and so no physical rights would be demanded for dispatching them. The generators would then receive $2p_3q_2 = \left(p_3 - \frac{\eta}{3}\right) q_2 + \left(p_3 + \frac{\eta}{3}\right) q_2$, as they should to provide the correct incentives. Of course, in general, such agreements among producers might raise concerns about collusive behavior, and this consideration may make bundling an unattractive policy option. It must also be the case that the ISO and the stakeholders share a common physical model of the network, so there is a match between what the ISO recognizes at "nets" and what the stakeholders can agree to do.

Consider now the *parsimonious* set of unidirectional physical rights. The number of physical rights from node 1 to node 2 is no longer a fixed number equal to K unlike in the case of an exhaustive set, but rather is determined *endogenously* by G_2's production. [This observation builds for our simple network on a more general point made by Chao and Peck (1996) in a perfectly competitive environment; this point has not always been well understood and certainly has not yet been fully incorporated into current reform proposals, and therefore is worth belaboring]. Because each unit of production in the South unloads the congested link by one

[46]"Netting" is called "counterscheduling" in the policy debate in California.

third the total number of rights available for bilateral trades between G_1 and the consumers should be equal to

$$K + \frac{q_2}{3},$$

resulting in the following constraint on production in the North:

$$\frac{q_1}{3} \leq K + \frac{q_2}{3}.$$

Furthermore the newly-created rights should be turned over to G_2 who then resells them at price η each to producers in the North. The total revenue for a unit production in the South is therefore $p_3 + \frac{\eta}{3}$, as it should be.

Note three potential difficulties with this arrangement: First, it would seem that bilateral trades between G_2 and consumers and the associated production in the South must be scheduled ahead of those in the North, so as to allow G_2 to resell the newly-created permits to generators in the North. This unfortunate sequentiality, which may disturb the price discovery process, might be circumvented by allowing G_2 to sell short (that is, to sell in advance) physical rights that it anticipates receiving at the scheduling date, with clearing and settlements occurring at that date.

Second, the use of a parsimonious set may face difficulties in situations in which a link may be constrained in opposite directions at different times of the day or seasons.

Third, one might worry about G_2 possessing market power in the physical rights market (besides that on the energy market). For the same reason as in the two-node network, G_2 may want to withhold some of the newly-created rights. To see this, let us distinguish between the number of rights, $q_2/3$, held by G_2 as a result of producing q_2, and the number of rights, $\hat{q}_2/3$, sold to generators in the North, where

$$\hat{q}_2 \leq q_2.$$

Production in the North is then

$$q_1 = \hat{q}_2 + 3K;$$

and because $p_3 = p_1 + \dfrac{\eta}{3}$, G_2's profit can be written as

$$p_3 q_2 - C_2(q_2) + \eta \frac{\hat{q}_2}{3}$$
$$= P\left(3K + q_2 + \hat{q}_2\right) q_2 - C_2(q_2) + \hat{q}_2 \left[P(3K + q_2 + \hat{q}_2) - C_1'(3K + \hat{q}_2)\right].$$

G_2 withholds none of the newly-created rights if and only if the derivative of its profit function with respect to $\hat{q}_2$ at $\hat{q}_2 = q_2$ is nonnegative, that is if and only if (using the first-order condition with respect to q_2)

$$C_2'(q_2) - C_1'(q_1) - q_2 C_1''(q_1) \geq 0.$$

As in the two-node network, G_2 trades off the need for substituting expensive for cheap power (which argues in favor of no withholding) and the desire to extract G_1's inframarginal rents (if any).

Finally, we note that an identical "withholding" strategy for G_2 is feasible under exhaustive rights and netting, as long as G_2 can choose to schedule some of its production in the South without netting it with an equal production in the North. Thus, *the exhaustive rights and netting do not differ with respect to their scope for withholding transmission capacity.* Similarly, prohibition of unmatched production by G_2 under exhaustive rights, or of withholding newly-created rights under parsimonious rights would be the counterpart to the capacity release program that we discussed above.

Observation #3: Closed-end physical rights portfolios.

Whichever way one proceeds, the thrust of the introduction of markets for physical rights is to have such rights traded among stakeholders. Efficiency requires that the rights corresponding to links with excess capacity be traded at zero price. But if such rights were indeed worthless, an investor or a stakeholder could costlessly create a spurious scarcity by purchasing a sufficient fraction of them and withholding some of them. The parties engaged in bilateral trades would then have to pay for more than one link.

Thus, it does not seem reasonable to organize separate markets for physical rights on the different links. Indeed stakeholders value *bundles* of rights, rather than individual rights (which per se are useless). In our context, this suggests that one could for example offer two bundles of rights. The first bundle, with K such rights, tailored for dispatching Northern production *on a stand-alone basis*, would give the rights to two units of capacity between nodes 1 and 3, and one unit between nodes 1 and 2 and between 2 and 3. The second bundle, tailored to *joint dispatching* of equal (netted) quantities at the two generation nodes gives no rights on the line from 1 to 2, and a unit right on lines 1 to 3 and 2 to 3. This approach has the benefit of preventing anyone from creating a spurious scarcity of rights on noncongested lines; more thought however should be devoted to the design of this portfolio of bundles in situations in which the location and the direction of the binding constraints is uncertain.

8.3 Loop flow: summing up

The extension to the three node network allows us to consider the effects of loop flow, an important and unique attribute of electric power networks. Loop flow creates complementarities between the behavior of generators at different nodes and, as a result, introduces a richer set of competitive interactions between generators than exist on a two-node network. Loop flow also allows for transmission congestion on different links and this further enriches the nature of competitive interactions that can arise on electric power networks. Nevertheless, the effects of transmission rights holding on market power on a three node network are conceptually similar to those on a two node network, taking into account the expanded set of competitive interactions. Finally, we have identified a number of institutional complexities that must be addressed to implement efficiently a physical rights system on a network with loop flow even in the absence of market power. If these institutional complexities are not addressed properly, players with market power may be able to further exploit these imperfections to their advantage.

9 Conclusion

We have demonstrated that when transmission rights are in positive net supply, their allocation can interact with pre-existing electricity seller or electricity buyer market power in ways that can enhance that market power, induce production inefficiency and reduce welfare. These effects are found on networks with or without loop flow. Whether and how transmission rights can have such effects depends upon the microstructure of the transmission rights market and the configuration of market power (location, buyer vs. seller) in the electricity market.

Both financial and physical transmission rights can enhance electricity seller or buyer market power in essentially the same ways. However, physical rights may potentially have worse welfare properties than financial rights since they can be withheld from the market, reducing effective transmission capacity and inducing production inefficiency. Appropriate rules requiring the release of unused rights (use-or-lose) and restrictions on the ability of generators with market power at the cheap node to capture the ex post value of these rights by buying electricity at the cheap node and reselling it at the expensive node (non-commitment) can improve the welfare properties of physical rights significantly. More generally, as restructured electricity sectors consider the creation and allocation of transmission rights, it is important that their potential adverse welfare effects be taken into account in the design of rights allocation mechanisms and regulatory rules governing the concentration of ownership and use of rights. Efforts to mitigate underlying electricity seller and buyer market power problems that appear to be endemic to electric power networks are obviously important as well.

Appendix : Welfare comparisons

To save on notation, let us assume that $C_2(q_2) = c_2 q_2$, that is, production in the South exhibits constant returns to scale (this is not essential). This assumption allows us to compare G_2's optimal price function when $K - q_1$ physical rights are withheld,

$$\hat{p}_2(q_1) \equiv \arg\max_{p_2} \left\{ p_2 \left[D(p_2) - q_1 \right] \; - C_2 \left(D(p_2) - q_1 \right) \right\},$$

with the price function that prevails when G_2 holds a fraction α_2 of financial rights (see our companion paper),

$$p_2(\alpha_2) \equiv \arg\max_{p_2} \left\{ p_2 \left[D(p_2) - (1 - \alpha_2) K \right] \; - C_2 \left(D(p_2) - K \right) \right\}.$$

Under constant returns in the South,

$$\hat{p}_2(q_1) = p_2 \left(1 - \frac{q_1}{K} \right).$$

Social welfare in all our variants is a simple, decreasing function of the price p_2 in the South and of the level of production, $K - q_1$, withheld in the North:

$$W \left(p_2 \, , \; K - q_1 \right) \equiv S \left(D(p_2) \right) - C_2 \left(D(p_2) - q_1 \right) - C_1(q_1),$$

where $S(\cdot)$ is the consumer gross surplus. Given local market power in the South, the constrained optimum is obtained when the price in the South is the monopoly price for the residual demand curve, $\hat{p}_2(K) = p_2(0)$, and when there is full production in the North ($q_1 = K$).

The upper bound, Π_1, for G_2's and the rights owners' joint profit under any institution is

$$\Pi_1 \equiv \max_{\{p_2 \, , \, q_1 \leq K\}} \left\{ p_2 D(p_2) - C_2 \left(D(p_2) - q_1 \right) - q_1 C_1'(q_1) \right\}.$$

This upper bound is obtained for $p_2 = \hat{p}_2(0)$ and $q_1 \leq K$ (with $q_1 < K$ if and only if $c_2 - C_1'(K) < K C_1''(K)$). Letting q_1^c ("c" for "commitment") denote the optimal q_1 in this program, let

$$W_1 \equiv W \left(\hat{p}_2(0) \, , \; K - q_1^c \right).$$

Let us also define

$$\Pi_2 \equiv \max_{p_2} \{p_2 D(p_2) - C_2 (D(p_2) - K) - K C_1'(K)\},$$

$$W_2 \equiv W(\hat{p}_2(0), 0),$$

$$\Pi_3 \equiv \max_{q_1} \{\hat{p}_2(q_1) D(\hat{p}_2(q_1)) - C_2 (D(\hat{p}_2(q_1) - q_1) - q_1 C_1'(q_1)\},$$

and letting q_1^{nc} ("nc" for "noncommitment") denote the optimal q_1 in the latter program,

$$W_3 \equiv W(\hat{p}_2(q_1^{nc}), K - q_1^{nc}).$$

Last, let

$$\Pi_4 \equiv \max_{p_2} \{p_2 [D(p_2) - K] - C_2 (D(p_2) - K) + [\hat{p}_2(K) - C_1'(K)] K\}.$$

$$\equiv \hat{p}_2(K) D(\hat{p}_2(K)) - C_2 (D(\hat{p}_2(K)) - K) - K C_1'(K),$$

and

$$W_4 \equiv W(\hat{p}_2(K), 0).$$

We have

$$\Pi_1 \geq \max(\Pi_2, \Pi_3) \geq \min(\Pi_2, \Pi_3) \geq \Pi_4,$$

$$W_4 > W_2 \geq W_1 \quad \text{and } W_4 > W_3.$$

We summarize the analyses of sections 3 and 4 in Figure 2.[47] Figure 2 assumes away free riding and therefore posits that gains from trade between the generator with market power and the rights owners are realized. In Figure 2, welfare decreases when moving east (increase in local market power) or north (increased withholdings).

[47]In the "no rights" case, the value of the ISO's merchandizing surplus, that is the value of the fictitious rights, is included in the measures of total profit and welfare. Alternatively, the "no rights" case stands for the situation in which G_2 is prevented by free riding or by regulation from buying the rights.

References

[1] Admati, A, Pfleiderer, P., and J. Zechner (1994) “Large Shareholder Activism, Risk Sharing and Financial Market Equilibrium,” *Journal of Political Economy*, 102: 1087-1130.

[2] Allaz, B. (1992) “Oligopoly, Uncertainty and Strategic Forward Transactions,” *International Journal of Industrial Organization*, 10(2):297–308.

[3] Allaz, B. and J.L. Vila (1993) “Cournot Competition, Forward Markets and Efficiency,” *Journal of Economic Theory*, 59(1):1–16.

[4] Borenstein, S., J. Bushnell and S. Stoft (1997) “The Competitive Effects of Transmission Capacity in a Deregulated Electricity Industry,” University of California Energy Institute.

[5] Burkart, M. , Gromb, D. and F. Panunzi (1998) ”Why Higher Takeover Premia Protect Minority Shareholders,” *Journal of Political Economy* ,106(1):172–204.

[6] Bushnell, J. (1999) “Transmission Rights and Market Power,” *Electricity Journal*, October: 77–85.

[7] Bushnell, J., and F. Wolak (1999) “Regulation and the Leverage of Local Market Power in the California Electricity Market,” Presented on the 1999 NBER Summer Industrial Organization Workshop, July.

[8] Chao, H.P., and S. Peck (1996) “A Market Mechanism for Electric Power Transmission,” *Journal of Regulatory Economics*, 10(1): 25–60.

[9] Green, R. (1992) “Contracts and the Pool,” mimeo, University of Cambridge, Department of Applied Economics.

[10] —– (1996) “The Electricity Contract Market,” mimeo, University of Cambridge, Department of Applied Economics.

[11] Green, R. and D. Newbery (1992) “Competition in the British Electricity Spot Market”, *Journal of Political Economy*, 100:929–953.

[12] Grossman, S. and O. Hart (1980) “Takeover Bids, the Free Rider Problem, and the Theory of the Corporation,” *Bell Journal of Economics* , 11: 42-64.

[13] Harvey, S., W.Hogan and S. Pope (1996) “Transmission Capacity Reservations Implemented Through a Spot market with Transmission Congestion Contracts,” *Electricity Journal* , 9(9): 42–55.

[14] Hogan, W. (1992) “Contract Networks for Electric Power Transmission,” *Journal of Regulatory Economics* , 4(3): 211–42.

[15] Holmström B. and B. Nalebuff (1992) “To the Raider Goes the Surplus? A Re-examination of the Free-Rider Problem,” *Journal of Economics and Management Strategy*, 1:37–62.

[16] Holmström, B. and J.Tirole (1998), “Private and Public Supply of Liquidity,” *Journal of Political Economy*, 106: 1-40.

[17] Joskow, P. (1996) “Introducing Competition into Regulated Network Industries: From Hierarchies to Markets in Electricity,” *Industrial and Corporate Change*, 5(2): 341–382.

[18] Joskow, P. and R. Schmalensee (1983), *Markets for Power*, MIT Press.

[19] Joskow, P. and J. Tirole (1999a) "Transmission Rights and Market Power on Electric Power Networks. I: Financial Rights," mimeo, MIT and IDEI.

[20] —– (1999b) "Transmission Rights and Market Power on Electric Power Networks. II: Physical Rights," mimeo, MIT and IDEI.

[21] Klemperer, P. and M. Meyer (1989) "Supply Function Equilibria in Oligopoly under Uncertainty," *Econometrica* 57: 1243–1277.

[22] Kyle, A., and J.L. Vila (1991) "Noise Trading and Takeovers," *Rand Journal of Economics*, 22: 54-71.

[23] Nasser, T.O. (1997) *Imperfect Markets for Power: Competition and Residual Regulation in the Electricity Industry* , unpublished PhD dissertation, MIT Department of Economics.

[24] Oren, S. (1996) "Preemption of TCCs and Deadweight Loss in Centrally Dispatched Electricity Systems with Competitive Generation," Power WP041, UC Berkeley.

[25] —– (1997) "Economic Inefficiency and Passive Transmission Rights in Congested Electricity Systems with Competitive Generation," *Energy Journal*, 18(1): 63–83.

[26] Oren, S., Spiller P., Varaiya, P. and F. Wu (1995) "Nodal Prices and Transmission Rights: A Critical Appraisal," *Electricity Journal*, 8: 24–35.

[27] Schweppe, F., Caramanis, M., Tabors, R. and R. Bohn (1988) *Spot Pricing of Electricity*, Kluwer Academic Publisher.

[28] Shleifer A. and R. Vishny (1986) "Greenmail, White Knights, and Shareholders' Interest," *Rand Journal of Economics*, 17: 293-309.

[29] Stoft, S. (1999) "Financial Transmission Rights Meet Cournot: How TCCs Curb Market Power," *The Energy Journal.*

[30] Tabors, R. (1996) "A Market-Based Approach to Transmission Pricing," *Electricity Journal*, 9(9): 61–67.

[31] Tirole, J. (1988) *The Theory of Industrial Organization*, Cambridge: MIT Press.

[32] Wolfram, C. (1999) "Measuring Duopoly Power in the British Electricity Spot Market," *American Economic Review*, 89(4): 805-826.